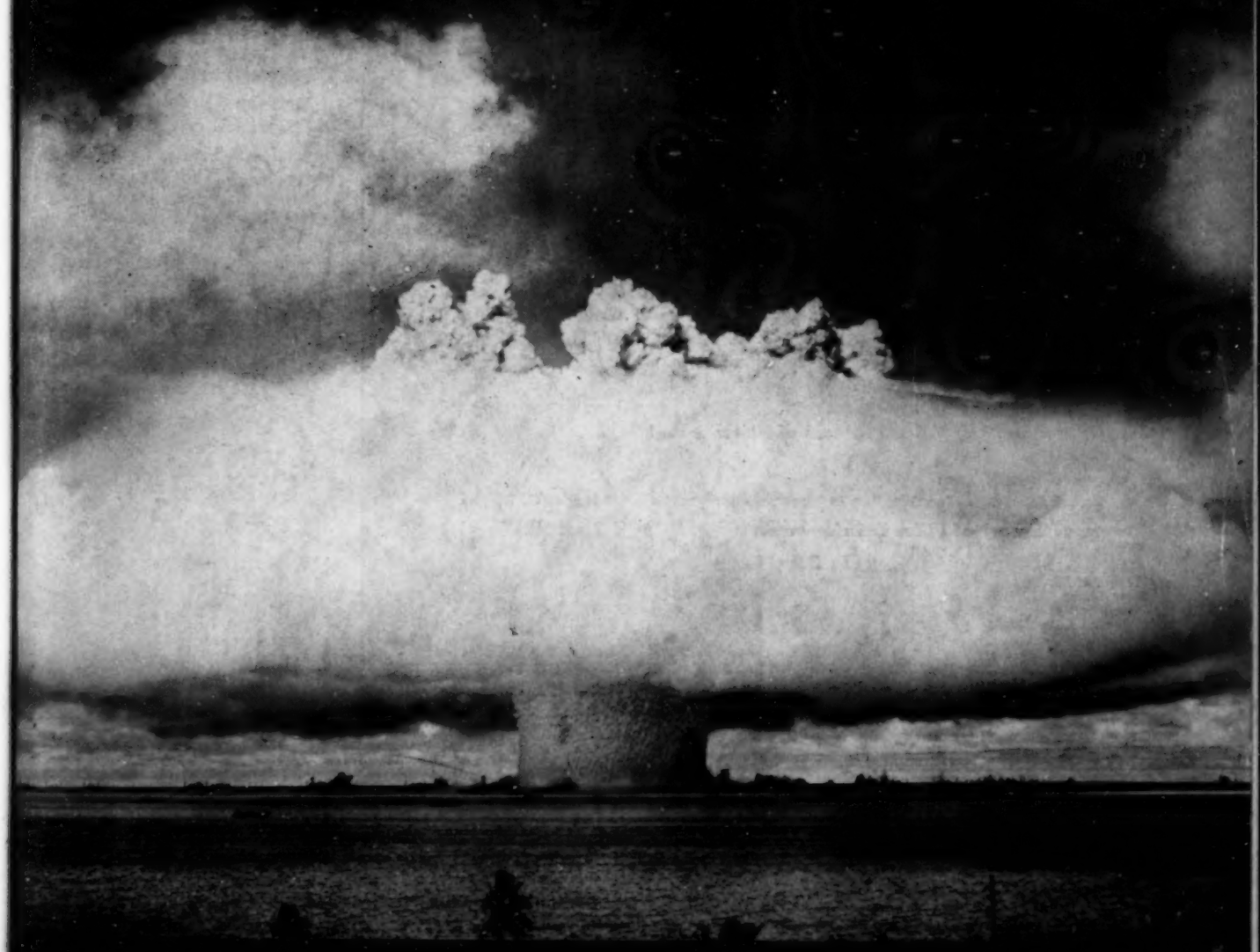
JOURNAL OF APPLIED PHYSICS



JOURNAL OF APPLIED PHYSICS

Vol. 18, No. 1 JANUARY, 1947

In This Issue

		Page
European Electron Induction Accelerators	HERMAN F. KAISER	1

Contributed Original Research

Ion Beams in High Voltage Tubes Using Differential Pumping	E. S. LAMAR AND W. W. BUECHNER	22
X-Ray Diffraction Studies of Color Variation of Iron Oxide Pigments	H. BIRNBAUM, H. COHEN, AND S. S. SIDHU	27
Mass Spectrometer for Leak Detection	A. O. NIER, C. M. STEVENS, A. HUSTRULID, AND T. A. ABBOTT	30
New Developments in Vacuum Engineering	ROBERT B. JACOBS AND HERBERT F. ZUHR	34
Magnetic Electron Microscope Objective: Contour Phenomena and the Attainment of High Resolving Power	JAMES HILLIER AND E. G. RAMBERG	48
Investigation of a Method for the Analysis of Smokes According to Particle Size	WILLIAM N. LIPSCOMB, T. R. RUBIN, AND J. H. STURDIVANT	72
Dislocation Theory as Applied by N.A.C.A. to the Creep of Metals	A. S. NOWICK AND E. S. MACHLIN	79
Application of the Sliding Thermocouple Method to the Determination of Temperatures at the Interface of a Moving Bullet and a Gun Barrel	E. L. ARMI, J. L. JOHNSON, R. C. MACHLER, AND N. E. POLSTER	88
Carbide Structures in Carburized Thoriated-Tungsten Filaments	C. W. HORSTING	95
Creep Deflections in Columns	JOSEPH MARIN	103
Effects of Magnetic Field on Oscillations and Noise in Hot-Cathode Arcs	J. D. COBINE AND C. J. GALLAGHER	110
Numerical Solution of Axisymmetrical Problems, with Applications to Electrostatics and Torsion	GEORGE SHORTLEY, ROYAL WELLER, PAUL DARBY, AND EDWARD H. GAMBLE	116
Note on a Boundary Value Problem of Reissner and Sagoci	IAN N. SNEDDON	130
Letters to the Editor:		
Determination of Magnification in Electron Micrographs	HENRY C. FROULA	19
Erratum: Theory of Filler Reinforcement	JANE M. DEWEY	132

In Every Issue

Here and There	18
New Books	20
New Booklets	21
Index to Advertisers	xxvi

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

Associate Editors

R. BOWLING BARNES
OTTO BEECK
W. F. BUSSE
SELIG HECHT
W. JAMES LYONS
PHILIP MORRISON
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND
JOHN VON NEUMANN

Publication of the American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*

MARGARET S. GRIFFIN, *Publications Manager*

THEODORE VORBURGER, *Business Manager*

JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD

PAUL E. KLOPSTEG, *Chairman*

WALLACE WATERFALL, *Secretary*

GEORGE B. PEGRAM, *Treasurer*

R. B. DOW
L. A. DUBRIDGE
W. F. FAIR, JR.
FLOYD A. FIRESTONE
HARVEY FLETCHER
R. C. GIBBS
K. S. GIBSON
W. H. MARKWOOD, JR.
A. H. PFUND
JOHN T. TATE
L. W. TAYLOR
A. G. WORTHING

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case School of Applied Science, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. AND CANADA	ELSEWHERE
To members and associate members of the American Institute of Physics	$5.00	$5.70
To all others	7.00	7.70

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 17, 1946—$108.50
Yearly back number rate: **$7.70**
Single copies: **$0.70** each.

Changes of address and proofs should be addressed to the Publications Manager.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1947, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 18, Number 1 January, 1947

European Electron Induction Accelerators

BY HERMAN F. KAISER

U. S. Naval Research Laboratory, Washington, D. C.

Betatrons, induction electron accelerators for the production of high energy *X*-radiation, were under development in Germany during the war and to a limited extent at present. This work is reviewed for the most part and details are given on the constructional features of 6- to 15-million volt betatrons and on the theory and design of 15- and 200-million volt betatrons. The smaller units, especially the Siemen's 6 Mv, are quite successful and more advanced than comparable American units. While no large machines were actually built, the projected 200-Mv Wideroe design introduces new features of value. Betatron patent ideas are reviewed.

I. HISTORY OF BETATRON DEVELOPMENT ON THE CONTINENT

THE first published work on electron induction acceleration to appear in Europe was the doctoral dissertation of Rolf Wideroe[1] of 1927 which was published as a paper[2] the following year. It is known, however, that a patent application for an electron induction accelerator was filed in Germany by J. Slepian (Westinghouse) soon after the filing of an application[3] in the United States in April, 1922. Because of a misunderstanding of the principles involved by a Dr. Smidt of the German Patent office who persisted in quoting the text of Abraham and Becker in support of his contention that it was impossible to both accelerate and guide an electron by a magnetic field at one and the same time, it was not granted until 1928. Whether or not Wideroe had knowledge of this patent is not known. At the same time experiments were being carried out by Breit and Tuve[4] at the Carnegie Institute in Washington which it is now known were correct in principle and might have been successful if pursued further. In 1929 a work by E. T. S. Walton,[5] following a suggestion by Lord, Rutherford, appeared in England. In this he discovered independently the flux conditions announced by Wideroe and made a further contribution on the theory of the magnetic fields required for assurance of a stable orbit. Successful experimental results unfortunately were not obtained.

Developments on the continent have been based on Wideroe's work and on the Steenbeck patents of 1937[6] in which the necessary conditions of Wideroe are set forth and supplemented

[1] Dissertation, Tech. Hochschule Aachen, 29 Oct. 1927.
[2] Archiv. f. Elektrotechnik **21**, 387 (1928).
[3] U. S. App. Ser. 548, 630, April 1, 1922, U. S. Pat. 1,645,304, Oct. 11, 1927.
[4] Carnegie Institution Yearbook **27** (1927–1928), p. 209.
[5] Proc. Camb. Phil. Soc. [9] **25**, 469 (Oct. 1929).
[6] DRP 656,378 and 698,867; U.S. Pat. 2,103,303, Dec. 8, 1937.

by the condition for producing axial stability by curvature of the guide field flux—a condition which was discovered independently in the United States by Breit and Tuve and later by Lawrence and Livingston for the cyclotron. An induction accelerator was built and tried by Steenbeck, who recorded production of 1.8-Mev radiation. Following Kerst's[7] publication of successful production of 2.5-Mev radiation at intensities of 15–100 milligrams of radiation, Steenbeck published a belated note on his work in 1943,[8] describing results obtained in 1935 and 1936.

The work of Kerst was a great incentive to the continental workers interested in this field, particularly Steenbeck and Wideroe. During the war the latter was the more active in this work, but Steenbeck set into action further work on this subject by Siemens Reiniger. Wideroe and the group that came to be associated with him in the wartime German betatron work were not in sympathy with the Nazi cause but were induced to continue such work for scientific and other considerations.

An unpublished investigation[9] interrupted by the war was carried out by James L. Tuck working in collaboration with L. Szilard at the Clarendon Laboratory in Oxford. It is now known that this investigation came very near to success.

The betatron development was carried out by three separate groups. One was the Megavolt Versuchsanstalt (MVA) or Megavolt Research Association which worked under the guidance of Dr. Wideroe. Siemens and the A.E.G. also had betatron projects. MVA laid plans for 15-, 30-, and a 200-Mv accelerators. The first was begun early in 1944 and brought to successful completion in the fall of the same year. The second was merely tentative. The design of the 200 Mv was prompted by news that the General Electric Company (U.S.A.) was constructing a 100-Mv betatron. The design of the 200-Mv accelerator was completed about this time and the construction work was turned over to the Brown-Boveri firm in Heidelberg. Occupation in Germany, however, came before any construction work could be begun.

[7] Phys. Rev. **60**, 47 (1941).
[8] Naturwiss. **19/20**, 234 (1943).
[9] Private communication from D. W. Kerst.

The basic ideas underlying the design of the 15- and 200-Mv induction accelerators were published by Wideroe in two papers.[10] These discuss the special problems involved such as the axial and radial stabilization by properly designed guiding fields, injection and extraction of electrons, the useful load of the apparatus, the maximum electron filling of the acceleration chamber, dependence on injection voltage, use of auxiliary lens systems especially for high radiation energies, the effects of gas molecules, technical transformer construction for high energy production and finally the fields of application of the betatron; cancer treatment, radiography, and nuclear physics.

In collaboration with the design work of Wideroe, a considerable amount of theoretical work was carried out by Touschek which was known to have been of invaluable aid in the development of the 15-Mv accelerator. Further theoretical work has been done by Touschek on the starting of electrons in the accelerator. Some of this work is along the lines initiated by Kerst and Serber[11] whose work was known to Touschek. Recent work of his, it is believed, may give an explanation of the discrepancy between the Kerst-Serber theory and experimental results.

The wartime betatron work of the Siemens Reiniger firm was based on the work of Steenbeck whose 1940 patent is mentioned above. An American license for its use by the General Electric Company was requested from Siemens shortly before the outbreak of war in 1941. The results published by Kerst[11] renewed an interest which had lagged and in 1942 (December) designs were under way for construction of an electron induction accelerator. Dr. Steenbeck gave the basic specifications which were incorporated into a design at Siemens and discussed in a conference on December 15, 1942 held together with Dr. Ganswindt, Dr. Kurt Bischoff, Dr. J. Patzeld, and Dr. Konrad Gund all of Siemens. At this time was also discussed a design of Dr. Gund in which he proposed an accelerator with a fixed magnetic guide field, surrounding an alternating magnetic field passing through a central core, a

[10] Wideroe, Part I, Archiv. f. Elektrotechnik **37**, 391–408 (1943). Part II remains in proof form.
[11] D. Kerst, Phys. Rev. **58**, 841 (1940); Phys. Rev. **59**, 110 (1941); and **60**, 47–53 (1941).

scheme originating from a paper of W. W. Jasinsky[12] in which the field-radius conditions are erroneously set forth. These are corrected in the design mentioned and several other ideas set forth. A patent application was filed by Siemens in July 1942[13] covering these and also premagnetization or bias field operation. Ideas for operation with biased fields were also evolved by Kerst at about this time.[14]

From the design worked out in the above conference and later meetings a successful 6-Mv betatron operating on 550 cycles was developed. A more detailed description of this unit will be given later.

II. ACTIVITIES OF THE MEGAVOLT RESEARCH ASSOCIATION (MVA)

A. Description of the 15-Mv Induction Electron Accelerator

General Considerations

Theoretical curves of the absorption coefficient as a function of electron energy have a minimum which lie between 3–10 Mev for most metals. A peak energy of 15 Mev was, therefore, chosen to bring the effective energy in the neighborhood of 7–8 Mev. (It has since been shown that 20 Mev satisfies these requirements better.) A theoretical sixfold increase in the range of radiographic thicknesses was expected. It was also desired to use the betatron for medical purposes but not too much was expected except for an improvement in the depth dose condition due to a lowered absorption coefficient. The desirability was recognized of using the betatron for the production of high energy electron beams in which the dissipation of energy by absorption in human tissue may be localized more effectively. This problem was under consideration but no practical results had been realized.

Constructional Details, Core

The radius of the electron equilibrium orbit was chosen as $r=14$ cm. The diameter of the induction pole was chosen as 220 mm (11-cm radius) and the air gap in the poles was made 2 cm. The average separation of the guide poles was chosen as 6 cm and the contour of the guide pole faces was determined from magnetic measurements and from measurements made by means of the electrolytic trough. This process was continued until the guide field had the properties required. The guide poles were so arranged with respect to the remainder of the apparatus that, as a result of the saturation of the guide field, the latter does not increase in proportion to the accelerating field from a certain point in the cycle onwards. This type of construction has some obvious advantages. The space between the acceleration pole and guide poles allows the use of flux biasing windings. It also has the advantage that no orbit expansion equipment is required, the electron striking the target upon saturation of the guide field and this without the attendant loss of energy which is present when the electron has to spiral into a smaller radius to strike a target. It may be noted, however, that this feature may still be obtained in Kersts' form of construction in which acceleration and guide fields arise from the same pole piece by a special reduction of the amount of iron at the edge of the pole piece. The approximate data for the main components (which are slightly above the initial estimates of Dr. Wideroe) are:

Iron weight	1000 kg	2200 lb.
Copper weight	200 kg	440 lb.
Total losses	2.5 kw at 50 cycles	
Dead load	250 kva	
Driving voltage	6 kv	

The details of core and pole assembly are shown in Fig. 1. The yoke is constructed from laminated iron of thickness equivalent to 29-gauge, 0.014-inch stock of approximately ½ watt per pound core loss (10,000 lines and 60 cycles). The lamination stack is not divided, i.e., no provision is made for cooling the core by means of air ducts. It was found that the heating was not excessive. The vertical members of the yoke, in contrast to American design, extend the full height of the apparatus, and the horizontal members are held between these by a special arrangement of clamping frames. This produces four vertical "shim" gaps. On the upper side these have means for adjustment of the height and level of the upper transverse member (yoke). This part carries the accelerating pole by means of six bolts circularly arranged. By means of a supporting ring, it also carries the guide pole and

[12] Archiv. f. Elektrotechnik **30**, 500 (1936).
[13] Siemens Akt. 151,465, VIII c/211g.
[14] D. W. Kerst, Phys. Rev. **68**, 233 (1945).

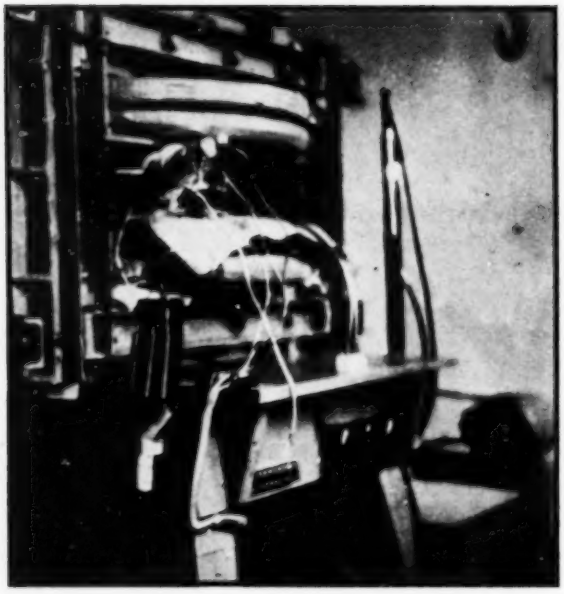


FIG. 1. View of 15-Mev betatron at Wrist.

coil boxes. The yoke laminations are clamped by transverse bolts through insulating tubes.

Acceleration Chamber

Glass acceleration chambers were made by the glass-blowing laboratory of the C. H. F. Müller firm. These have an approximately elliptical cross section. The inside walls are coated with graphite or silver coatings and the tubes have either two or three radial openings with stems on the periphery for evacuation or for mounting the electron injection assembly. Vacuum was produced by an oil diffusion pump backed by a three-stage mercury diffusion pump with a liquid air trap separating the two. The mercury diffusion pump was backed in turn by a rotary oil pump. This unnecessary complication of pumps was caused by the scarcity of available and usable equipment.

The vacuum was measured by means of a Phillips vacuum meter placed between the two diffusion pumps and by means of an ionization gauge sealed to the acceleration chamber on the injection stem. It was noted that operation was not possible at pressures above 10^{-5} mm Hg and steady yields were obtained only at 10^{-6}-mm pressure.

The electron gun or injector is shown in Fig. 2 and follows in principle the design introduced by Kerst. It consists of a tungsten filament (oxide coated electron sources were also used) surrounded by a semi-cylindrical Wehnelt or Faraday cage and of two anode plates insulated from each other. The cage serves, on application of a potential between itself and the filament, to concentrate the electron stream. The split anode, on the two halves of which different potentials could be applied has, ignoring for a moment the necessary acceleration of electrons, a purpose similar to that of the deflection plates of a cathode-ray tube. It is thus possible by applying various potentials to these plates to give the electrons various initial projection angles in the plane of the orbit. The necessary correction of the electron beam in the vertical direction is obtained by a small rotation of the whole injector system or stem on its ground glass joint. A small bar of tungsten is fastened to the innermost anode plate with its long axis vertical to serve as a target for x-ray production.

The filament is heated continuously. The injection voltage is applied only for a short time at the beginning of the acceleration period by means of a special arrangement to be described below, partly to charge the entire injector structure to as high a voltage as possible and partly to avoid filling the electron accelerator chamber with stray electrons.

Connections

Figure 3 shows the entire diagram of connections. Alternating current of 6 kv is generated by a regulating transformer and supplied to the driving circuit in which the dead load of the circuit is compensated by a capacitor bank, the

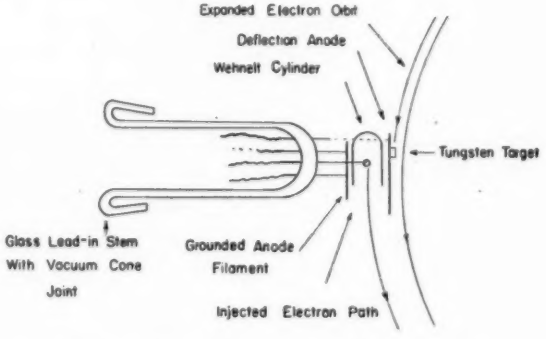


FIG. 2. Electron injector gun for 15-Mev betatron (schematic).

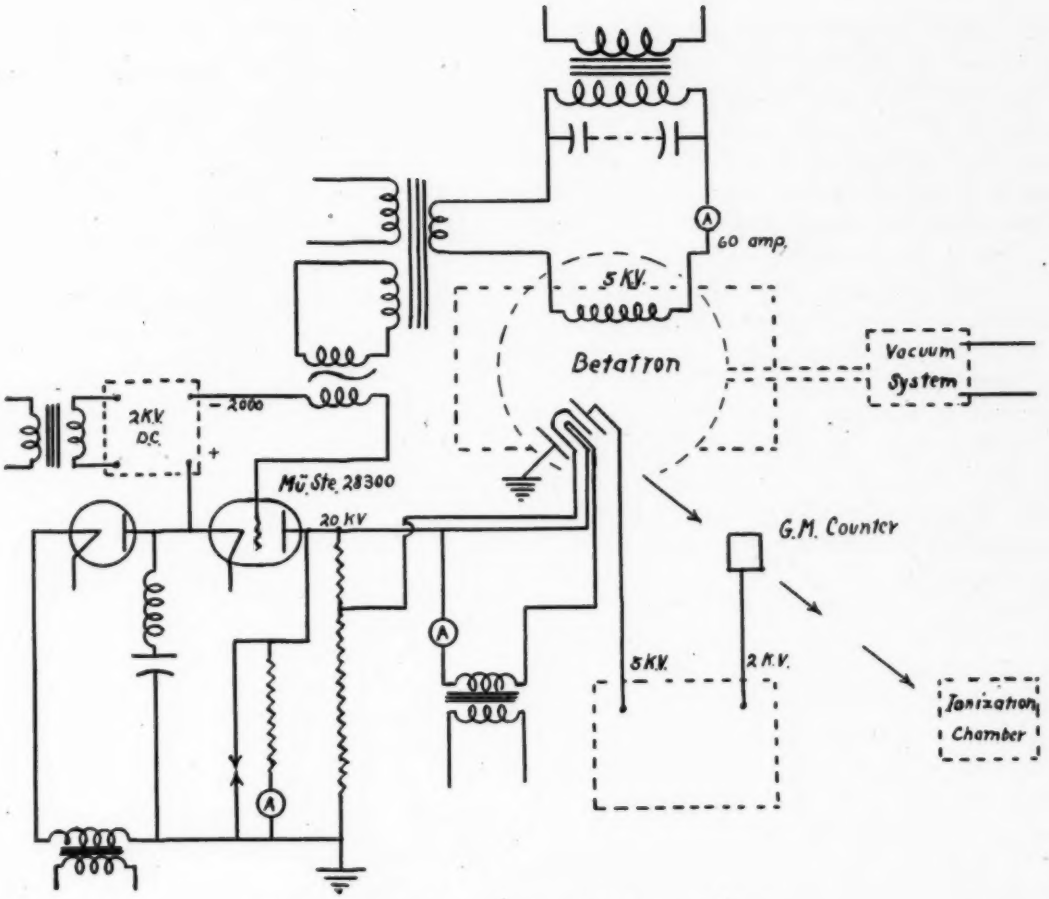


FIG. 3. Diagram of connections for 15-Mev betatron.

capacity of which may be varied between limits. The injection voltage is produced by H.V. equipment from which the potential may be applied at the proper instant near the beginning of the acceleration period by means of a special grid controlled high voltage rectifier tube. The driving circuit of the betatron passes through a small utility transformer which in turn operates a peaking transformer for the production of a sharp positive pulse to remove momentarily the large negative potential on the grid of the control valve. In earlier experiments, a mechanical contactor driven by a synchronous motor was used to apply the tripping pulse but this gave much trouble due to fouling of contacts. The circuit diagram also shows a 5-kv constant potential circuit used for applying the deflection voltages to the split anode, the innermost plate of which is connected together with the acceleration chamber coating to ground.

Operation

In the early stages of operation radiation is detected by means of a counter tube which of course becomes useless when the machine is brought into adjustment and radiation of considerable intensity is produced. Ionization chambers are then used and depended upon for measurement of the radiation intensity. With shielding the counter tube may still be used for indicating the point in the cycle at which radiation is emitted directly upon an oscilloscope in which the horizontal or time base is supplied by the a.c. driving voltage.

In operation the vacuum is checked, the driving circuit is energized, and the phase of the injection pulse is determined and corrected by means of the phase shifting device. On starting the counter and cathode-ray tube, a radiation output is generally observed. By rotation of the ground glass joint of the injector, by application

of a suitable deflector voltage, by changing the position of the acceleration chamber in the machine, and also by adjustment of the air gap in the guide field it is possible to gradually bring about an optimum yield. Then by keeping these variables constant quite reproducible results could be depended upon which were then only influenced by the vacuum conditions.

To check the quality of the radiation produced, various methods were resorted to: (a) observation of the output radiation pulse on the oscilloscope allowed an estimate of the time of acceleration of the electrons and a deduction of their maximum energy which in general was around 13–15 Mev, (b) absorption measurements on the radiation which, however, are somewhat uncertain because of the shallow minimum in the absorption *versus* energy curve, (c) rapid dying out of radiation intensity on reducing the amplitude of the exciting current, and (d) production of artificial radioactivity by irradiation of various materials. These phenomena are especially pronounced in the case of copper and silver.

After the expected quality of radiation had been established, attention was directed to improving the intensity of the radiation. It was hopéd to attain a maximum electron filling of the acceleration chamber such as could be expected from the following formula deduced by Wideroe for the effect of space charge*:

$$i=\frac{4\pi n^3(\Delta r)^3(1-n)f^2rB_{\max}}{c^2(-dr)}\text{ amp.}$$

Δr: radial oscillation amplitude,
dr: change in Δr,

in which r is the radius of the orbit, n the exponent in field-radio relation, f the frequency, $B_{\max}$ the maximum guide field strength, and the other quantities their usual significance. This formula indicates a value which is usually about ten times that observed experimentally. In operation yields ranging from 30 grams to 1 kilogram radium equivalent were obtained. It was believed that this discrepancy is caused by the release of gases by bombardment of the walls of the acceleration chamber and that better results could be obtained only by use of baked out and sealed-off acceleration chambers.

* A brief derivation is given in Appendix I.

Program for Further Development

The program of the Megavolt Versuchsanstalt as described in June 1945 was: (a) further work to increase the radiation yield of the 15-Mev betatron and (b) to build as soon as possible a duplicate of this equipment for use in medical research to allow the first unit to be used solely for technical development. This was, of course, the wartime program. At present scientific research is strictly limited by the military government of the occupational forces.

Under (a) it was planned to make further experiments on acceleration chambers fitted with electron lens systems which might, in view of the greater stabilization forces made available by their use, bring about an increase in radiation output intensity. It was also planned to study the tangential electron injection technique of Wideroe once more because of the possible advantage accruing from the absence of an injector structure in the acceleration chamber; also to study the problem of releasing a beam of electrons from the apparatus, a subject of considerable importance.

Another intention was the design and construction of a 30-Mev unit for use in nuclear research. This project was to have been undertaken at an early date and some of the design calculations had been made along with those for the 15-Mev betatron.

It was felt that an apparatus of equal size but operating on the premagnetization scheme of Wideroe would be useful in testing such questions as to whether or not the final energy of the accelerator would be affected by the radiation damping as predicted by Iwanenko and Pomerantschuk.[15]

A final project was the design and construction of a 200-Mev accelerator using the premagnetization scheme. This will be described below.

B. Description of the 200-Mev Induction Electron Accelerator

Preliminary Considerations

The design of the 200-Mev betatron was the outgrowth of some preliminary design work carried out by Wideroe for a 100-Mev unit.[10] This was based on his experience with the con-

[15] Phys. Rev. 65, 43 (1944).

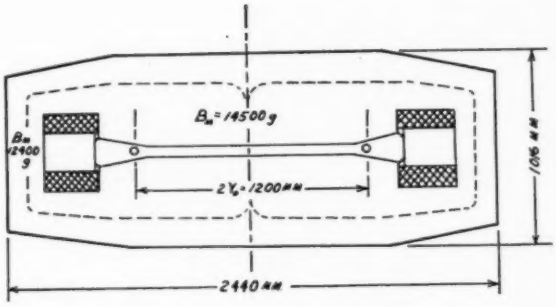


FIG. 4. Design for a 100-Mev betatron.

struction of large transformers by the Brown-Boveri firm with which he was associated. An economical design for transformer construction used by this firm consists of a radial construction in which the radial laminations are welded at the top and bottom. In this type of construction an exactly circularly shaped iron section is obtained and the required guide pole shapes are easily designed and adapted. The necessary cooling ducts are likewise easily incorporated into the core. The minimum amount of iron is used if the yoke is continuous as in a shell transformer. This lends itself to a rugged construction having a small yoke height and large cooling surface. It also offers a certain amount of radiation protection by shielding the acceleration chamber, access to which is obtained by apertures cut into the yoke.

A preliminary design for a 100-Mev unit is shown in Fig. 4. Here yoke, core, and iron return form a common iron unit constructed from radial laminations. By removal of the ten-ton upper half, access is had to the windings and the acceleration chamber. The core is crossed with air ducts to remove the 13.6-kw iron losses (0.35-mm laminations with a loss of 1.1 w/kg are assumed). The maximum induction of the core is 14,500 gauss for a stacking factor of 77 percent. The yoke and iron returns are able to dissipate 28 kw by surface cooling. In these parts the stacking factor is taken as 90 percent and the maximum inductions chosen are 14,500 and 12,500 gauss, respectively. The total weight is 20 tons.

The air gap for an equilibrium orbit of radius $=60$ cm is 5 cm and about 20,000 effective ampere turns are required to produce in it a guide field of 5600 gauss. To accomodate the exciting winding (46.5 kg of copper, assuming a current density of 2 amperes/mm^2), a window of 800 cm^2 is provided. Thus a space factor 12.4 percent is obtained, allowing insulation of the winding to 30–50 kv. This voltage is so chosen that it will be the most economical one considering cost of capacitors. The copper losses of the transformer will be 5–6 kw and the required dead load 7500 kva. The purely constructional costs aside from development and research are estimated at 80,000 RM for the betatron, 120,000 RM for the capacitors, thus making a total cost of 200,000 RM. Using this 100-Mev design as a basis the weights, costs, dimensions, etc. of still larger machines may be estimated as has been done by Wideroe. Of interest here is the estimation for a 200-Mev machine which would have the following characteristics: core diameter 4.89 meters, core height 2.04 meters, weight 160 tons, core losses 400 kw, radiation load 400 kw, and cost about 1,120,000 RM. It is apparent that unless further refinement were introduced betatrons of still higher energy would be extremely costly.

Improvement of Yield by Means of Auxillary Lenses

It was the hope of Wideroe and his group to increase the electron filling of the acceleration chamber by means of electro-static or electromagnetic lenses. In the 100-Mev unit it was planned to fill the air gap as completely as possible with the acceleration chamber and to provide multiple injectors and guide fields on its periphery. Allowing four guide fields for each oscillation of the electron, 28 cathodes will be required for seven oscillations per revolution. With an average electron beam of 1 ma ($V_i=300$ kv) the frequency of the injection voltage will be 17,000 cycles for a 100-Mev betatron operating at 60 cycles. This gives a charging period of 2×10^{-5} second per cycle. The average charging current for these times must therefore be one ampere and requires, assuming an efficiency of 10 percent, a total cathode emission surface of 500 mm^2. For 28 cathodes this amounts to 18 mm^2 per cathode. Cathode windows will also have to be of this order of size. To prevent any electronic charges from building up on the acceleration chamber the outside must be neutralized and grounded.

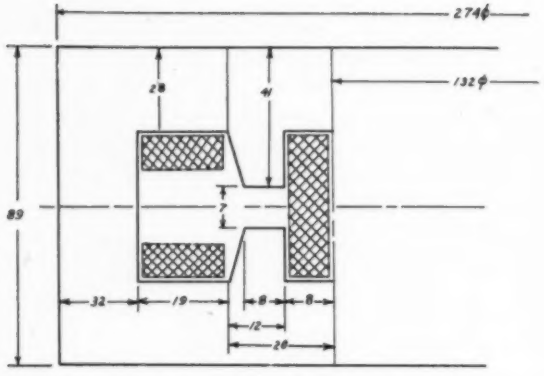


FIG. 5. Schematic construction of 200-Mev betatron (Wideroe).

*Projected Design and General Description of 200-Mev Betatron**

As mentioned above it was evidently necessary to build a 200-Mev betatron with further refinement of design to keep the size and weight within reason. A first step in this direction was the elimination of the air gap of the 100-Mev design. Even with this feature the diameter of the machine would still be about three meters but a very considerable reduction in the weight was possible. Late in 1944 ideas had been crystallized and the Brown-Boveri firm which is skilled in the construction of large transformers was contracted in August. By mid-October a first rough design had been drawn by the firm which was discussed at length in Heidelberg. In this conference the Brown-Boveri firm was represented by its director Dr. Meyer-Delius, Chief Construction Engineer Otto Weiss, and Dr. Helmut Boecker, detail designer. The Megavolt Versuchsanstalt was represented by Dr. Wideroe and Dr. Kollath. The construction of models and a small experimental unit for the study of special problems was also considered at this time. Exact calculations were to be completed before the end of the year. In November, the MVA received some questions and designs to which reply was made by letter. In January, the first plans for construction were complete and the MVA was informed but it was too late for another conference and discussion of these plans. Since that time, the MVA had no contact with BBC and all work on this project came to an end with the occupation of Germany early in 1944.

The conference held at Heidelberg evolved a design that had the following general characteristics, which may be seen in Figs. 5 and 6. A radially laminated core of 1320 mm diameter was to be surrounded by ring-shaped radially laminated parts of about 200 mm width. These carry the guide poles which have an average separation of 7 cm and also the compensating windings. About these are placed the main exciter windings, divided above and below the median plane. To complete the magnetic induction circuit four mantels are provided 90 degrees apart on the circumference. A difficulty arising from this feature will be discussed below. The preliminary problems of the original design included the purely mechanical construction of the laminated parts of the machine, the fastening of the guide poles to the machine, the division of the transformer to facilitate installation of the acceleration chamber, pumping and electrical connections, and adjustments to the guide poles. It was planned to purchase about 40 tons of 0.35-mm transformer sheet (29 gauge) having a core loss of about 1.1 watts per kilogram.

The upper pole shoes were to be capable of being raised, and the four mantel yokes were to be displaced radially to allow access for leads, etc. A special difficulty was in the evacuation arrangements and in the cooling of the exciter windings. It was proposed to mount the inner winding directly adjacent to the core in a stationary oil case.

Final Form Decided Upon

The general features of the final form evolved in the design work are shown in Fig. 6. In vertical section is shown the central core with its cooling system consisting of a number of parallel helically shaped cooling ducts winding through the body of the core and connecting at the upper and lower ends of the core to the compressed air inflow and return ducts 5. The core has a stacking factor of about 70 percent. By means of ledges about its periphery, a ring of radially laminated pieces making up the pole shoes is held in proper position, and so that the upper pole shoe ring may be

* More detailed design calculations of Dr. Wideroe for both the 15- and 200-Mev units may be found in the author's Naval Technical Mission T.R. 331–45. (Now released to Dept., Commerce Publications Board.)

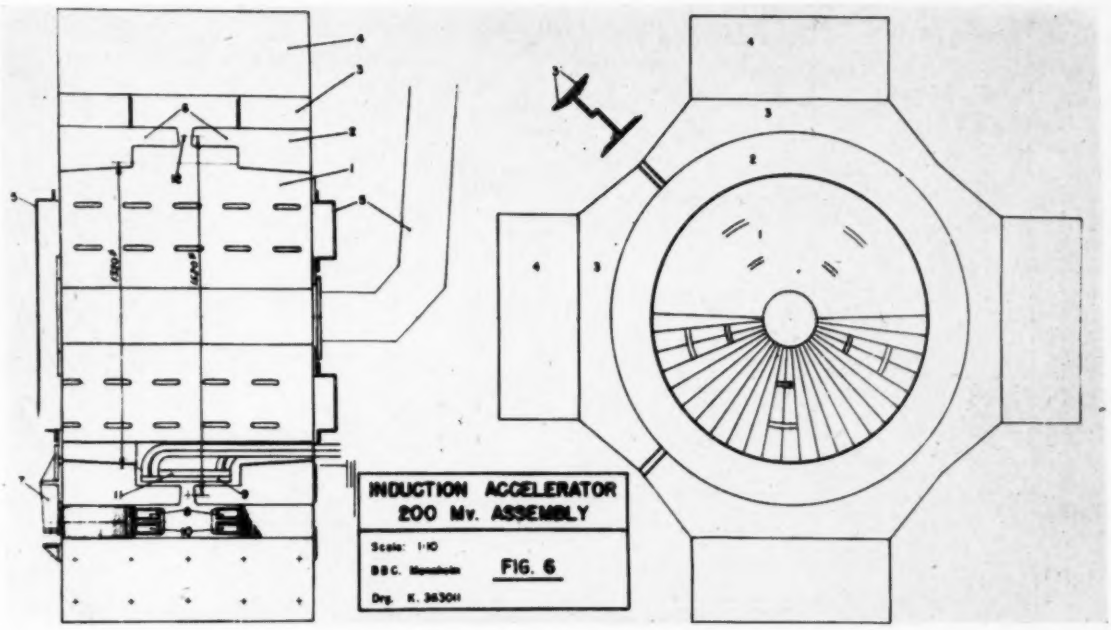


FIG. 6. Assembly of 200-Mev betatron.

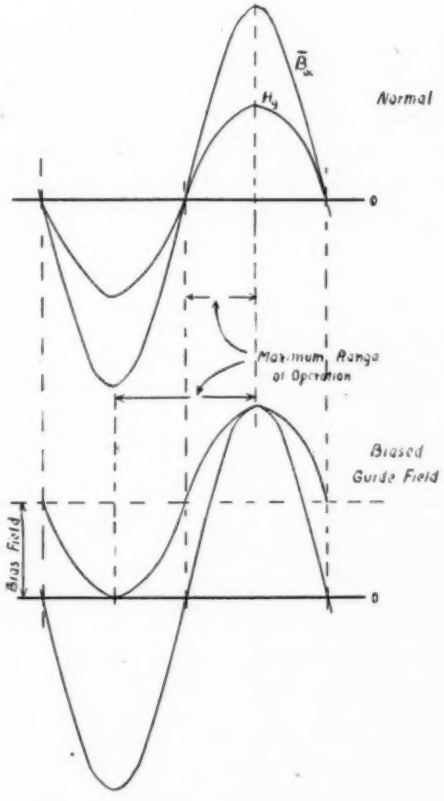


FIG. 7. Principle of operation with biased guide field.

raised together with the connecting ring *3*. This carries on its lower side the main a.c. exciter winding and the d.c. premagnetization winding. The original intention to complete the gapless magnetic circuit by means of the four mantels led to the difficulty that the uniformity of the magnetic flux in the guide poles will be disturbed by the locking of the mantels to the guide poles at only a few points. To avoid this difficulty, it was proposed to add an additional ring *3* (horizontally laminated) which serves the purpose of smoothing the flux from the core into the mantels *4*. This is electrically broken between adjacent mantels with a zig-zag joint to avoid an electrically continuous circuit about the flux passing to the guide poles. The figure shows the principal and compensating windings *8* and *9*, and the direct-current premagnetizing windings *10* and *11*.

By elimination of the air gap in this design a large central accelerating flux is produced by the main windings. These windings also serve to produce the guide field which at all times has to bear a fixed relation to the induction flux in order to maintain a stable electron orbit. In this design it is furthermore proposed to premagnetize the guide field so that it will be operative over nearly

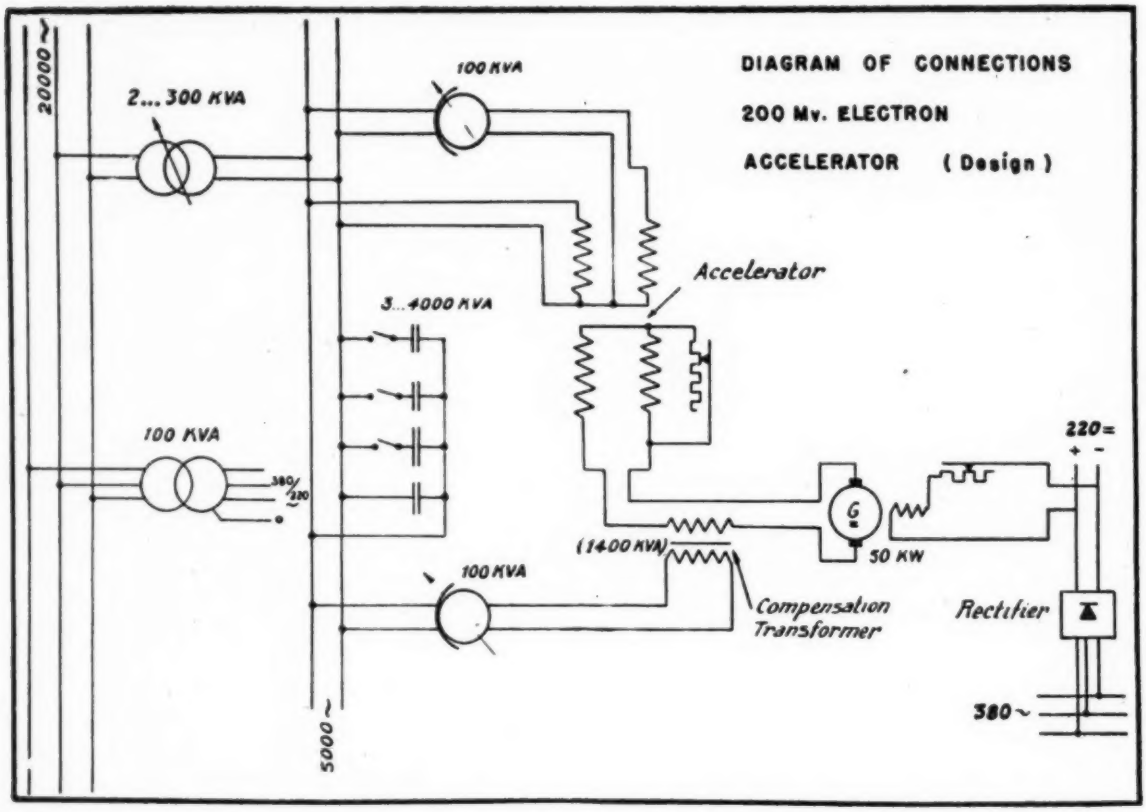


FIG. 8.

a half a cycle instead of a quarter of a cycle, thus permitting the induction field to accelerate the electrons for more than a quarter of a cycle. This principle is illustrated in Fig. 7, and has recently been discussed in the literature.[14,16]

The air gap of the earlier design provided the necessary relationship between the induction and guide fields by insuring their proportionality. Since the use of it leads to unnecessarily large power requirements, an attempt is made to provide the necessary relationship by a winding in the inner coil space (compensating winding) on which is impressed a fixed potential of the corresponding magnitude. For practical purposes the main magnetizing winding and its compensating windings may be connected to the same driving source and the flux *vs.* induction relationship may be regulated by proper proportioning of the windings of the two coils. For fine adjustment a regulation transformer may also be connected into the circuit. Figure 8 shows the manner in which these windings are to be energized from the power sources and the necessary compensating transformer connection to neutralize the a.c. voltage induced in the d.c. windings.

For this design the following windings were planned (see Fig. 6):

1. An alternating current winding placed outside of the guide pole.
2. A compensating winding correspondingly placed about the induction pole to depress the a.c. magnetization of the core to such an extent that the average induction of the core is only twice that of the induction at the center of the guide pole (fulfillment of the necessary 1:2 relation).
3. A premagnetization winding for the guide pole of which the induction is to vary from zero to a maximum.
4. A compensating winding corresponding to the above which, while in principle is not necessary, is used to prevent an undue loading of the machine by a d.c. magnetization with undesirably high induction and saturation characteristics.

Further general electrical characteristics on the design are:

[16] W. F. Westendorp, J. App. Phys. **16**, 657 (1945).

Total induced flux	105×10^6 maxwells
Induction in core	15,000 gauss
Induction in yoke	13,900 gauss
Induction in mantels (4)	12,200 gauss
Induction in guide pole	15,000 gauss
Total iron weight	26 tons
Volts/turn at equilibrium orbit	366 volts
Maximum electron energy	200.5 Mev
Operating power, driving circuit	7900 kva
Operating power, compensating windings	4800 kva
Resulting net load	3100 kva
Iron losses about	67 kw.

Details on the exact form of the guide poles were not determined these being a feature to be worked out during the construction.

Stabilization of Orbit and Improvement of Orbit Control

The magnetic stabilization forces, brought about by the suitable shaping of the guide poles, are known to diminish with an increase in size of the betatron while retaining the same sectional dimension of the acceleration chamber. Wideroe and Touschek have made a theoretical study of the additional stabilization that may be achieved by electron optical lenses. These may, in their simplest form, consist of a set of series connected coils spaced closely around the circumference of the acceleration chamber, the axis of each being tangent to the orbit. The principal advantage of such a lens system is that a greater electron filling of the acceleration chamber should be possible.

The maximum beam current is, in theory, determined by the effect of the mutual repulsion due to Coulomb forces of the electrons. These forces are compensated again by the magnetic forces mentioned above. Excess electrons thus injected are driven out of the beam and end on the walls of the acceleration chamber. The repulsive forces, although decreasing with time after injection of the electrons, are still relatively large. The effectiveness of the electron lens will be greatest in the early stages of the injection process. Wideroe has estimated that, in the 200-Mev design, it is possible to increase the stabilization forces for 10-kv injection about 1350-fold and that this would in turn increase the beam current from 0.1 to 135.0 microamperes. In the 15-Mev betatron the stabilization forces are greater and the relative improvement to be expected is much less. The improvement factor is here about 45 which would increase the current from 0.5 to 22.5 microamperes.

The magnetic lens system was tested out on the 15-Mev betatron at Wrist without any marked success. Further work on this scheme nevertheless is contemplated.

A further subject of concern in the design of the 200-Mev machine was the possibility of energy losses in the machine due to radiation damping set forth by Iwanenko and Pomerantschuk[16] and by others recently.[17,18] Views of Touschek and Lens are that as a result of the radiation damping a radiation of frequency 10^7–10^8 cycles is generated. This radiation, behaving like visible radiation, might in part be reflected by the metallized walls of the acceleration chamber and be conserved. Certainty on this point however did not prevail and it was hoped to study such effects experimentally on a smaller machine before completion of the large 200-Mev betatron. It was also believed that a possible remedy lay in increasing the driving frequency considerably.

III. BETATRON DEVELOPMENTS OF SIEMENS-REINIGER

Mention has been made previously in this paper of the activities of Dr. Steenbeck and his associates at Siemens-Reiniger. This firm at Erlangen has under development three units two of 5–7-Mev energy and one of 15 Mev. Only the former two are in advanced stages. Of these, the one operating on 550-cycle power is in operation; the other which is to operate on 50 cycles and embody some new ideas in the electron acceleration chamber is still in construction. The 550-cycle unit works quite well and has a radiation output equivalent to that produced by 12–20 grams of radium. The output depends very much on the vacuum conditions and hopeful extrapolation of yields based on the vacuum pressures in use lead to a belief in a possible 50-gram output. The Siemens firm was contemplating putting this machine into production. The betatron is shown in Fig. 9.

[17] J. P. Blewett, Phys. Rev. **69**, 87 (1946).
[18] L. I. Schiff, Rev. Sci. Inst. **17**, 6 (1946).

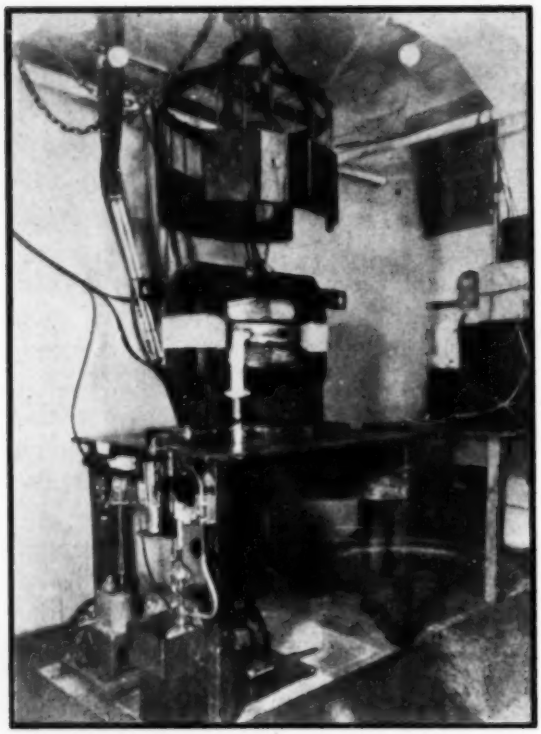


FIG. 9. View of Siemens-Reiniger 6-Mev betatron.

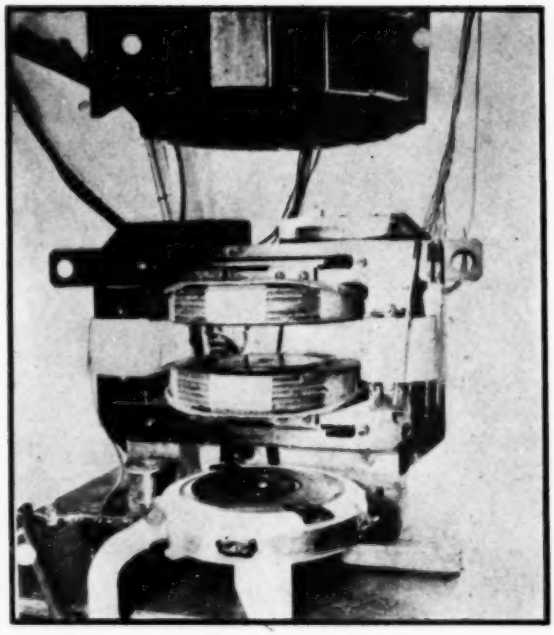


FIG. 10. View of 6-Mev betatron with acceleration chamber removed.

General description of 6-Mev, 550-cycle betatron

This equipment, to be described in more detail below, follows closely the betatron of Kerst in general principles although the constructional details are somewhat different and ingenious. The electrons are injected into the acceleration chamber by a single electron gun which is similar to that of Kerst but modified by the use of auxillary electric deflection electrodes. Radically different methods of electron injection are contained in patent applications but these schemes were mostly in the design stage. A point of departure from the Kerst model lies in the fact that the electrons are injected near the center of the apparatus rather than on the periphery and also that the electrons are injected in that portion of the field between the guide poles where the magnetization forces of stabilization are either weak or absent. After injection into the chamber, the electrons proceed in spiral orbits into a final equilibrium orbit of nearly fixed radius. X-rays are produced after acceleration in the latter orbit by orbit expansion produced by momentary introduction of current into a central winding which increases the flux in the induction pole or may weaken the flux in the guide field.

Magnetic Yoke and Windings

The general features of the yoke, exciter windings, and the acceleration chamber are shown in Figs. 10–13. As shown in Fig. 11 there is a yoke made of bolted stacks of laminations to make an approximately rectangular form with an open center to accommodate the coils and pole pieces. It may be noted that: (a) no gaps are provided for air cooling, (b) no radial construction is used in the large pole pieces which are made approximately circular by the expedient of making parallel lamination packets of different lengths, and (c) that the "staggering" method is employed to improve the general shape of the yoke. Cooling of the main guide poles is largely taken care of by the water cooled exciter windings which are hollow brass tubes on which special Litz wire is wound. Cooling of the injector mount is effected by air passing through ducts parallel to the induction pole and communicating with a suction (or pressure) duct at the base of the machine.

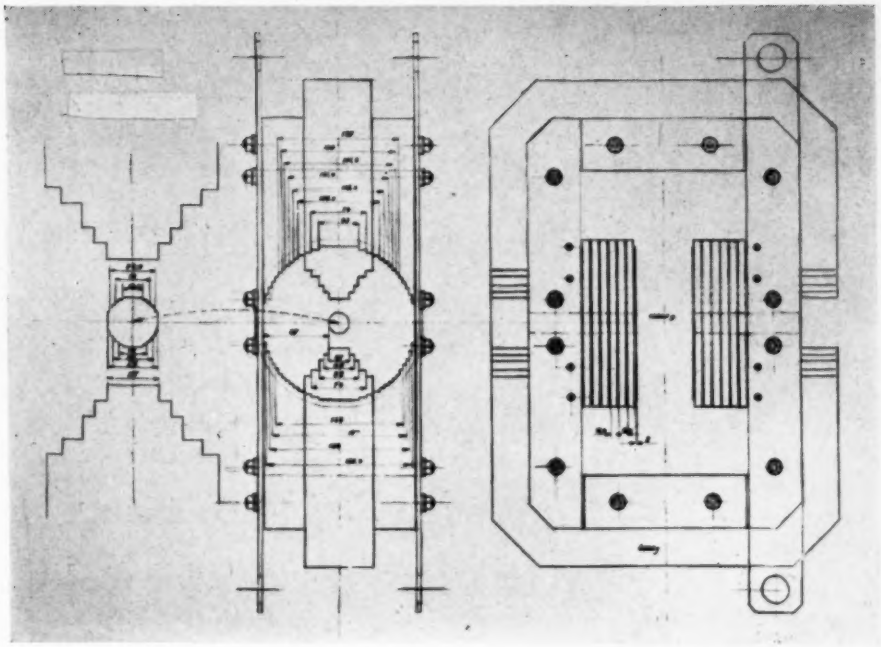


FIG. 11. Photograph of drawing showing yoke and construction of 6-Mev betatron. Since figures are not easily legible, the following scale may be helpful. As printed, reduction is $\frac{1}{10}$.

The guide pole pieces are finished by machining to precise flatness and against these rest the radially laminated packets which are built into the covers of the acceleration chamber and give the guide field its proper shape and intensity distribution. The major portion of the pole piece is thus built integral with the yoke and requires no special arrangement for supporting. At the center of the pole pieces may be seen a hole of approximately circular cross section extending continuously from the bottom to the top of the yoke. It connects at the lower side to the air supply duct, centers the middle section of the acceleration chamber and provides access from the upper side to screws for clamping or tightening the acceleration chamber assembly in place.

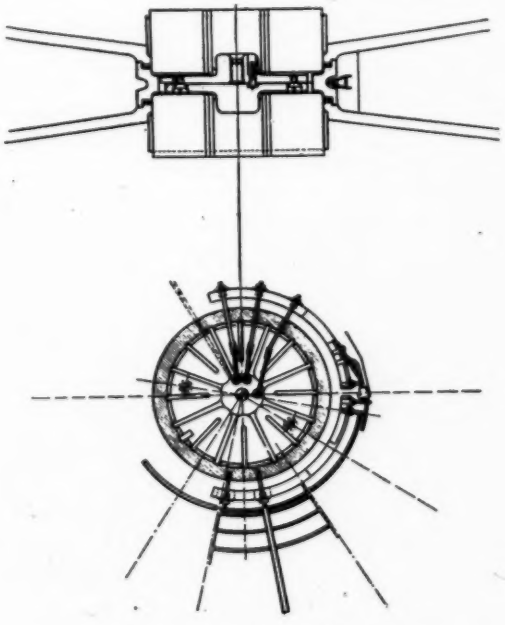


FIG. 12. Electron injection system for Siemens 6-Mev betatron.

The windings consist of 6-mm brass tubing which is wound transversely with five parallel strands of 5 sq. mm total section Litz wire, or of 9 parallel strands of somewhat smaller section Litz wire. These tubing turns were covered with tape, varnished, and formed into coils of 36 turns each (six high six deep) mounted on coil forms which are supported at the sides of the yoke by the binding plates of the yoke.

Acceleration Chamber

The acceleration chamber, shown in Figs. 10 and 12, is the most ingenious and intricate part of the equipment. It is constructed from a porcelain-like material called "Kallait" and made so that it is completely demountable as may be noted in one of the figures. Essentially, it is a body ring with side exhaust port and with circular

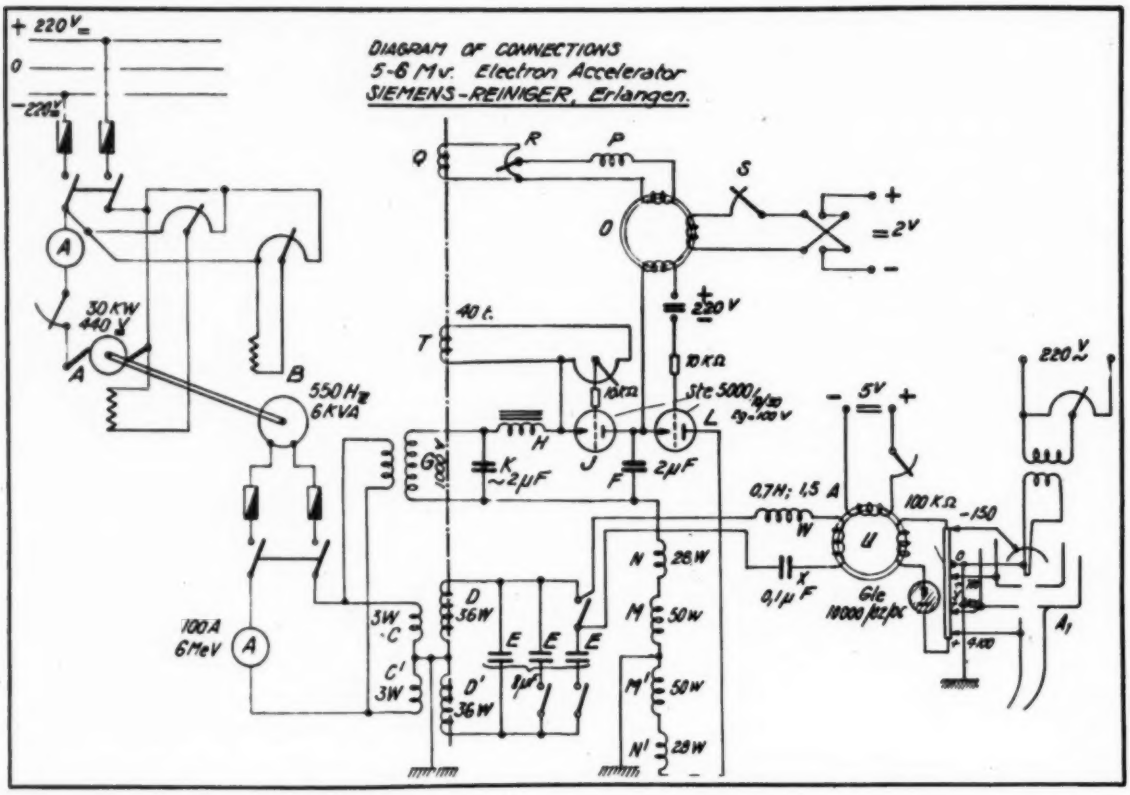


FIG. 13. Diagram of connections, 6-Mev betatron.

grooves for the gasket sealing lids which form the top and bottom of the acceleration chamber. These lids have built into their outward faces the radially laminated iron "faces" which on the outside are flat to match exactly the flat machined guide pole stump and on the inside are shaped to give the proper field shape and intensity distribution to the guide field. Near the center is located a cylindrical well into which is fitted the winding for producing orbit expansion by increasing the induction flux and weakening the guide field. The two lids mentioned have a central opening, the edges of which also make a gasket seat on the central "Kallait" core piece, which furthermore serves for carrying the electron injector structure and also all the necessary electrical lead-ins for its operation. It is cooled as noted above by an air stream passing through the center aperture of the yoke. Another feature of the acceleration chamber is its electrostatic shielding. This is accomplished by metallizing the inner walls of the top and bottom in grid fashion to break up complete electrical circuits about the pole or in its faces. Sections of this metallic coating are connected to radial leads distributed about the periphery of the wall ring. On the outside, these leads are connected together with high resistances and one point is grounded. The body ring also has a metallic spray coating which is contacted by the leads to the coating. On one side of the chamber is the target which is a small rectangular bar of tungsten held with its axis parallel to the field and which is mounted on one of the lead-in stems. On this side of the chamber a slot is provided in the wall of the body ring over which is cemented a thin metal foil window to minimize the absorption of x-rays and electrons passing out from the target. The direction of the radiation beam is approximately 45° to the long axis of the yoke, passing out of the betatron at a point just forward of the pump stem as shown in Fig. 9. This type of construction while apparently complicated is well suited to experimental work especially in view of necessary alterations to test out projected ideas on electron injection and field control.

Circuit Connections

Figure 13 shows the circuit diagram for the 550-cycle betatron. Power is obtained from a 550-cycle motor generator unit consisting of a 440-volt, 30-kw, direct-current motor driving a 6-kva alternator. Frequency is regulated by a manual speed control of the motor and no equipment is used to stabilize the speed. The alternator is a special low voltage machine supplying 100 amperes for 6-Mev operation. This current is passed through six primary turns on the betatron which may be seen passing over the main driving coils in Fig. 10. These are divided into two sets and center grounded, likewise the main windings. Their 36-turn terminals are connected directly to a capacitor bank of 8-mf capacity. A part of the parallel capacitor connection is used to operate a peaking transformer used to provide the voltage pulse for the electron injection. A 5-volt, d.c. winding provides means for shifting the phase of the injection. The output pulse of about 10 kv is passed into a 100,000-ohm potential divider from which various voltages applied to the injector electrodes are obtained. The filament is heated by a separate transformer and is grounded together with one point of the potential divider.

The orbit expansion circuit derives power from a 1-kv utility transformer connected in parallel to the primary coil of the machine. The output passes over a smoothing filter, K, and H, and a grid-controlled rectifier which controls the charging of the 2-mf capacitor F. The grid-control feature of the rectification is not indispensable. Control is obtained from a forty-turn winding T divided on the two legs of the machine. A similar winding Q supplies power to another peaking transformer O having a phase-shifting bias winding. This actuates the thyratron L (L and J are Osram Ste 5000/10/30). The capacitor T is discharged at the desired operating point through the two fifty-turn, orbit expansion coils M and M'. In series with these are two 28-turn, compensating winding N and N' wound about the legs of the yoke whose purpose is to compensate the magnetization effect of the orbit expansion current pulse in the yoke. It may be noted that the injection and orbit expansion pulses are phased properly by having one peaker transformer actuated by the primary driving current and the other by the voltage derived from the primary flux which is in quadrature therewith.

Review of Betatron Patent Ideas of the Siemens-Reiniger and C. H. F. Müller Firms

While the actual electron accelerator developments of Siemens-Reiniger seem to have covered only patent claims which are fairly well represented in one application[19] entered July 28, 1942, a number of ideas are contained in their present work and in other patent applications which will be reviewed briefly.

In an interesting experimental device use is made of a thermo element on the tungsten target of the betatron to measure the actual electronic current which is converted into radiation and heat. This is a more efficient indicator of output than the electron current passing from chamber to ground.

The use of the positive ion plasma for the purpose of increasing the electron density in the acceleration orbit by virtue of the electronic focusing and stabilization resulting from the positive ion space charge has been considered in a semi-theoretical way. This topic is on the program for further study.

Ideas contained in patent applications include the following:

a. Construction of the magnetic acceleration chamber in such a way that the magnetic guide field poles form the vacuum chamber itself and even contain the magnetizing windings to produce the guide field. This idea has met with experimental difficulties. It was intended for an accelerator for producing fairly strong circulating currents with several hundred kv energies.

b. Use of a built-in saturable magnetic core to keep the guide field constant for a short time during the injection period to increase the available injection time.

c. In relation to (b) use of a constant injection voltage during time in which guide field is constant.

d. Electron injection radially and normal to the field with asymptotic approach to an equilibrium orbit together with mechanism for transfer of orbit to a stable acceleration region or orbit.

e. A ring cathode with auxillary radial electric accelerating field. Applicable only on certain conditions.

f. Polygonal or elliptical ring cathodes to overcome limitations of (e).

g. A 1:2 magnetic field principle operative in asymptotic injection.

---

[19] Siemens Akt. 151,565, VIIIc/211g.

h. Time dependant guide field further characterized by pre-magnetization of a special magnetic core with a.c. in quadrature with the guide field.

i. Electron source consisting of a pair of oxide emittor plates, above, below, and parallel to median plane of accelerator. An acceleration chamber with an open center.

j. Accelerator as in (i) in which a plurality of electron orbits pass through the axis and produce there a high electron concentration.

k. Acceleration in a stable orbit with either a time variant or time invariant guide fields, together with an electrostatic arrangement for release of an electron beam from the orbit.

l. Orbit expansion by means of local disturbance of guide fields by special windings located at advantageous points or areas in field.

m. Arrangements for d.c. premagnetized guide field, also time variant guide fields.

n. Production of injection potentials by special windings in induction field.

Patent applications[20,21] credited to Dr. Müller of the C. H. F. Müller firm also contain the idea of using a ring cathode and radial electric acceleration fields mentioned in Sections (d)–(f) above. These contain perhaps the novelty of using magnetic shielding material to protect the filament of the cathode from magnetic forces and also have a departure in the proposal: first, to accelerate the electrons by induction within an inner portion of the guide field, then to pass these on by means of an applied radial electric field to an outer region where they may be further accelerated in a stable orbit. In the course of the passage, these are to pass through a retarding radial electric field region, the purpose of which is to bring the radial speed of the accelerated electron back to values that will allow them to be introduced into the equilibrium orbit. Electrons that are slow radially naturally form a closed circular path tending to return them to the cathode. A difficulty presents itself in the delaying field in that it may draw electrons oscillating radially about the equilibrium orbit back into this space. To avoid this difficulty, it is proposed to inject the electron in a line parallel to but above or below the median or orbital plane. This, it is believed, will cause the electron captured by the equilibrium orbit to execute spiral tracks which will make difficult the reentry of electrons into the decelerating radial electric field.

[20] C. H. F. Müller M 159,065 VIII 10/21g.
[21] C. H. F. Müller M 159,741 VIII c/21g (43/44).

In another Müller's application[22] the idea of spiral orbits about an equilibrium orbit is again set forth, this time as a possible means of avoiding collisions of injected electrons with the injector, the spiraling being so controlled that the electron will miss the injector. This is to be accomplished by applying a circular magnetic field whose lines of force are parallel to the electron orbit itself which field is to be applied only during the beginning of the injection.

In an application[23] it is proposed to release electrons after acceleration in a fixed equilibrium orbit in such a way that, on expansion of the orbit by saturation of the magnetic field, the electrons will move outward into a field varying so rapidly with distance that the spacing between successive turns of the spiral electron orbits becomes quite large. It then will be likely that an electron may strike a target or a window covering a certain angular range in its last revolution. No solution is given as to how the electron is so be properly phased to arrive at the desired instant in the required range.

Another application[24] contains an idea similar to that contained in Part (c) of the Siemen's applications. In this the magnetic guide field or the accelerating field is to be kept constant for a certain short time interval during the electron injection process by introducing into the electrical exciting circuit some controlling rectifier valves. These are usually a pair in parallel, one of which "saturates" in a certain time of operation after application of potential and thus produces the momentary constancy of field required. After a desired time the second grid controlled valve is fired and carries the current as a conductor of low resistance for the remainder of the half cycle. The scheme may be extended to produce a similar effect on the negative half-cycle. A C. H. F. Müller application[25] describes an electron injector in which the equilibrium orbit is allowed to pass through the injector structure without obstruction. The application includes the use of a manifold of such units about the accelerator orbit.

A very late application[26] describes an iron-less

[22] C. H. F. Müller M 159,714, VIII c/21g (1/3/44).
[23] C. H. F. Müller M 189,257, VIII c/21g (28/12/43).
[24] C. H. F. Müller/Dr. Müller 2000 (date unknown).
[25] C. H. F. Müller Appl. A 3224 of 25/9/44.
[26] C. H. F. Müller, Dr. Mü/Ro/160 Jan. 26, 1945.

betatron. The scheme, in which inductions considerably greater than 10,000 gauss are possible, is to consist of a coil of few turns and an acceleration tube of relatively small dimensions. The coil is of a hemispherical shape to give proper direction to the field lines and proper distribution of the flux through the orbit in the acceleration chamber. Because of the high power requirements due to lack of iron the betatron is to be operated by a sudden discharge of electricity stored in a capacitor bank. It is also proposed to obtain desired circuit time constants by proper design of self-inductance or by inclusion of external circuit impedance. The application further includes possible use of iron in such parts of the equipment where fields of 10,000 gauss are to be used. Iron, however, is not to be used in the regions of most intense fields, that is, 20,000 to 100,000 gauss.

IV. SUMMARY AND CONCLUSION

The principal betatron developments which have been carried out during the war by several German firms or organizations have been discussed above in some detail. In the way of evaluation of these developments, especially in comparison with developments which have proceeded simultaneously in the U.S.A., the following notes may be added:

The developments seem to have the same basis in preceding scientific literatures; the main reference sources are the works of Wideroe and the patents of Steenbeck and of Kerst, upon which developments agreeing fairly closely in general principle have been made. The success of these developments has been of various degrees. The works of Wideroe and the associated groups MVA and BBC have resulted in an operating 15-Mev electron accelerator described above, and some projected designs. The x-ray output of this unit, reported to be as high as 1 kg of radium equivalent and as low as 30 grams at various times of experimental work, compares fairly well taking the upper figure with early American 20-Mev units allowance being made for difference in maximum energy. The technique of electron injection, although subjected to theoretical investigation, lacks considerably in refinement, perhaps partly due to the difficulties encountered in carrying on experimental work at all. The betatron of Wideroe presents a somewhat different design from that of Kerst in that saturating guide poles are used. This feature, however, has opened up some design possibilities which could not be foreseen in the Kerst design in which the guide fields are produced on the same pole unit carrying the accelerating flux. A comparison in regard to weight indicates perhaps a slightly more efficient use of iron which consideration however is influenced considerably by the dimensions of the air gaps.

The most interesting and remarkable development due to the Wideroe group has been the design of a 200-Mev unit in which a very considerable economy of iron may be found. An interesting comparison may be made in the way of weight, between this design which is to weigh about 40 tons for a 200-Mev rating and the 130 tons of the American General Electric Company 100-Mev machine. In this 200-Mev design certain improvements, such as the use of premagnetization, are considered. At present it may be premature and optimistic to believe that all of these improvements will be practically realized on a large betatron until such have been demonstrated on a smaller scale.

The most successful betatron development in Germany is perhaps that of the Siemens-Reiniger 6-Mev, 550-cycle betatron. This unit has a good output of radiation for its small size and weight and operates with almost unbelievable steady output. The actual details of electron injection follow American practice to some degree with the difference that inside injection is employed. Some interesting differences lies in the electron starting. X-ray production occurs in larger American machines by orbit expansion but somewhat different orbit expansion arrangements are used. The most interesting feature is the adaptability of the unit to various arbitrary injection and acceleration experiments. American betatron practice has advanced to the use of sealed off porcelain acceleration chambers. The use of porcelain, however, was anticipated by Steenbeck who used a tube of such material in 1935.

Although no visit could be made to Berlin, it is understood that the AEG were engaged in betatron development also and that their work is based largely on the preliminary achievements of Wideroe and Steenbeck.

In connection with the developments reviewed above there are to be noted many interesting and ingenious technical details. A review of the patent applications indicates a wealth of ideas which have been systematically studied only in part.

APPENDIX I

Considering both the force of the electrostatic field of the beam of electrons in the equilibrium orbit, and the magnetic attraction force between this beam and the test electron moving parallel to it at a distance Δr their difference will be seen to be

$$\Delta P_1 = \frac{eQc^2}{\pi r_0 \Delta r}(1-\beta^2), \quad Q: \text{charge in orbit.}$$

The difference in centrifugal and magnetic forces acting on the electron for a small deviation Δr from the equilibrium orbit r_0 is

$$\Delta P_2 = \frac{e^2B^2}{m}(1-n)\frac{\Delta r}{r_0}, \quad n: \text{field law index}$$

to a first approximation. Equating these differences and solving for Q the total charge in the equilibrium orbit is

$$Q = \frac{\pi(1-n)(\Delta r)^2(r_0B)^2}{r_0E_0(1-\beta^2)}, \quad E_0 = m_0c^2/e$$

$$= \frac{\pi(1-n)}{r_0c^2}(\Delta r)^2(E^2+2EE_0)\frac{(E+E_0)^2}{E_0^2}.$$

Eliminating E by means of the damping relation: $-dr/r = \frac{1}{2}dE/E$ from which using $dE = rvdB$ and $dB \sim \omega B_m \cdot 2\pi r/v$

$$E = \frac{2\pi^2r^2\Delta r B_m \cdot f}{-dr} \quad \text{(injection voltage)}$$

f: operating frequency

we have finally that the average current supplied $i = f \cdot Q$ is

$$i = \frac{4\pi^3(\Delta r)^3(1-n)f^2rB_m}{-dr \cdot c^2}.$$

Here and There

Fiber Society—High-Polymer Joint Meeting

The third meeting of the Division of High-Polymer Physics, which was exclusively devoted to textile and fiber research, was held in Maury Hall of the University of Virginia on September 26–28, 1946. The meeting was sponsored jointly with the Fiber Society. Approximately 105 members of the Division and the Fiber Society and guests registered at the meeting.

The program was arranged by a joint committee under the co-chairmanship of Lewis Larrick and W. A. Sisson. Dr. Larrick also headed the local committee, in which undertaking he was ably and enthusiastically assisted by the staff of the Institute of Textile Technology.

In addition to addresses of welcome by Ivy F. Lewis, Dean, University of Virginia, and Ward Delaney, President, Institute of Textile Technology, three addresses of non-technical nature were given in the course of the three-day meeting. P. Larose of the National Research Council in Ottawa spoke on "Textile Research in Canada." Carl M. Conrad delivered a paper commemorating the life and work of the late Enoch Karrer, Fellow of the American Physical Society and a pioneer in high-polymer research. H. Wickliffe Rose presented, with numerous slides, impressions of Japan and the Japanese textile industry gained on a recent trip there.

Thirteen technical papers were presented. Titles and authors appear hereunder.

W. James Lyons
Division Secretary

"Air and Water Vapor Permeability of Fabrics," L. W. Rainard, E. L. Wilson, and A. L. Nestmann, Institute of Textile Technology.
"Moisture Transfer to and from Tire Cords Encased in GR-S," W. James Lyons,* Hilda M. Ziifle, Mary L. Nelson, and Trinidad Mares, Southern Regional Research Laboratory.
"Curl in Woven Textile Fabrics," Ross Whitman, Kendall Mills.
"Tricot," C. W. Bendigo, *Textile World*.
"Tensile and Torsional Properties of Textile Fibers," L. G. Ray, Jr., E. I. du Pont de Nemours and Company.
"The Creep Behavior of Textile Fibers," M. T. O'Shaughnessy,† E. I. du Pont de Nemours and Company.
"Electron Microscopical Studies of Natural Cellulose Fibers," William G. Kinsinger and Charles W. Hock, Hercules Powder Company.
"The Measurement of Irregular Distances and its Use in Studying the Swelling of Cotton Fibers," Enoch Karrer and R. S. Orr, Southern Regional Research Laboratory.
"Performance of Tensile Testing Machines," Herbert F. Schiefer and Josephine M. Blandford, National Bureau of Standards.
"Apparatus for Determining the Coefficient of Friction of Running Yard," Francis B. Breazeale, The American Enka Corporation.
"Spun Nylon," L. L. Larson, E. I. du Pont de Nemours and Company.
"An Investigation of the Load-Deformation Characteristics of Celanese and Wool with Particular Reference to the Use of These Materials in Carpet Pile," R. W. Work, A. M. Thorne, and M. R. Livingston, Celanese Corporation of America.
"The Design of Parachute Shroud Lines as Influenced by Fundamental Energy Absorption Properties," E. R. Kaswell, Fabric Research Laboratories.

* Resigned April 1945.
† Now at Brooklyn Polytechnic Institute.

New Appointments

Three new appointments to the staff of the National Bureau of Standards were recently announced. Dr. B. J. Miller is chief of the recently organized Guided Missile Electronics Section, Dr. H. K. Skramstad is chief of the Guided Missiles Section, and Dr. Ladislaus L. Marton is principal physicist in the Electronics Section.

Awards

Professor P. W. Bridgman of the Physics Laboratories at Harvard University has been awarded the Nobel prize in physics for 1946.

Dr. Willis R. Whitney, organizer and first director of the General Electric Company's Research Laboratory, has received the Industrial Research Institute Medal for his pioneer work.

Mathematical Tables Progress

The Mathematical Tables Project, 150 Nassau Street, New York, New York, in its progress report for October stated that work had been completed under the following headings:

Table of spherical Bessel functions, Volume II,
"Manila rope" distribution problem,
Spherical scattering functions for complex arguments,
Computations related to hydraulic analogy of shock wave intersections.

Information on Preparation of Graphs

The editor of *American Journal of Physics* has announced that anyone wishing to have a reprint of the article "Preparation of Graphs for Physical Papers," which appeared in the March–April 1946 issue of that journal, may receive one without charge by addressing *American Journal of Physics*, Wabash College, Crawfordsville, Indiana. The article suggests "practices intended for the guidance of authors and draftsmen in preparing graphs for reproduction in technical publications." It further states that "most of the recommendations are in accord with those recently made by a committee of the American Standards Association."

Postdoctoral Fellowship for Women

Announcement will be made early in March of the award of the third Sigma Delta Epsilon Fellowship of $1500. Women with the equivalent of a Ph.D. degree, carrying on research in the mathematical, physical, or biological sciences, who need financial assistance and give evidence of high ability and promise, are eligible. During the term of her appointment the appointee must devote the major part of her time to the approved research project, and not engage in other work for remuneration (unless such work shall have received the written approval of the Board before the award of the fellowship). Application blanks may be obtained from Dr. Louise S. McDowell, 28 Dover Road, Wellesley, Boston 81, Massachusetts. Applications should be submitted before February 1, 1947.

High-Polymer Lecture Series

During the academic year 1946–1947 a series of lectures on the properties of high polymers is being given at the National Bureau of Standards in Washington. They have been arranged by Robert Simha. The lecturers include the nation's leading scientists in this field from industry and universities. The lectures are open to the public without charge and are held from 7 to 9 o'clock in the evening in Room 214 of the Chemistry Building, National Bureau of Standards. Remaining dates in the course are January 22 and 30, February 27, March 6 and 28, April 24, May 8 and 29, June 5 and 12.

Necrology

Dr. Wilbur B. Rayton, director of the Scientific Bureau of Bausch and Lomb Optical Company, died October 31. He had been with the company since 1908 and was appointed head of the Bureau in 1926. Two months before his death he had received the Navy Ordnance Development Award.

Calendar of Meetings

January

15–17	American Society of Civil Engineers, New York, New York
27–30	American Society of Heating and Ventilating Engineers, New York, New York
27–31	American Institute of Electrical Engineers, New York, New York
28–30	Institute of Aeronautical Sciences, New York, New York
30–Feb. 1	American Physical Society, American Association of Physics Teachers, and Division of High-Polymer Physics (American Physical Society) (Joint Meeting), New York, New York

February

15	American Geophysical Union, Pasadena, California (Section of Hydrology)
16–19	American Institute of Chemical Engineers, Louisville, Kentucky (Regional Meeting)
20–22	Optical Society of America, New York, New York
22	American Mathematical Society, New York, New York
24–25	Inter-Society Color Council, New York, New York (Joint Meeting with the Technical Association of the Pulp and Paper Industry on the 25th)
24–27	Technical Association of the Pulp and Paper Industry, New York, New York
24–28	American Society for Testing Materials, Philadelphia, Pennsylvania

Letters to the Editor

Determination of Magnification in Electron Micrographs

HENRY C. FROULA
Assistant Physicist, Armour Research Foundation of Illinois Institute of Technology, Chicago, Illinois
(Received November 1, 1946)

Electron microscopes are often calibrated through the use of a grating replica as a specimen. The assignment of magnifications to subsequent micrographs then depends,

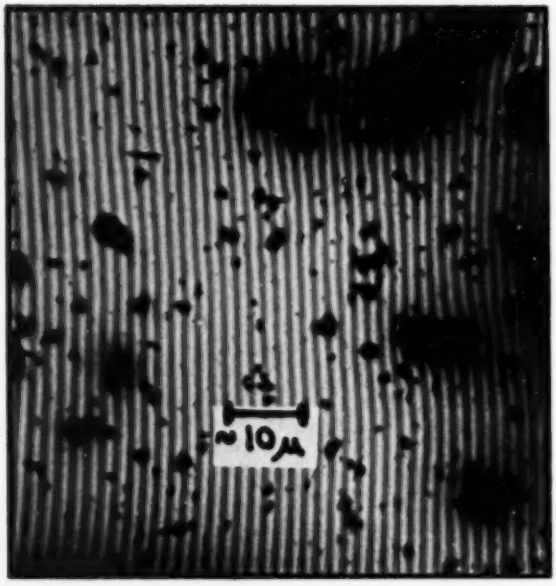


FIG. 1. An electron micrograph of coal dust supported by a 15,000-line per inch grating replica, prepared by the second method described herein. Magnification, approximately 1000×.

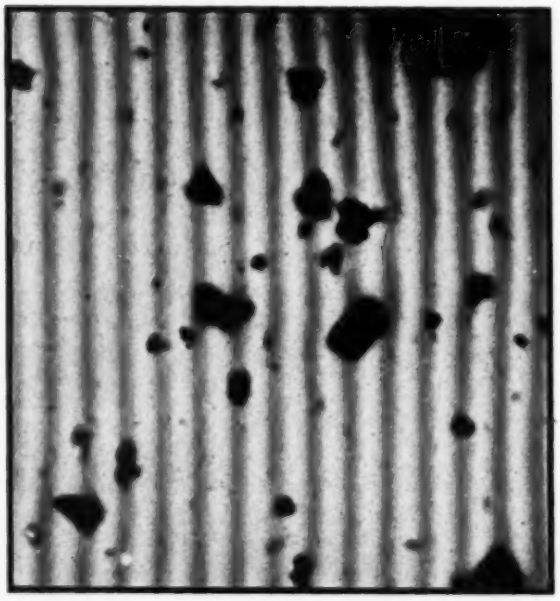


FIG. 2. A limited part of the field of Fig. 1 at a greater magnification, approximately 3000×.

essentially, either upon the constancy of the original calibration or upon a knowledge of how it varies.

Difficulties and skepticism in regard to the translation from the original calibration to subsequent micrographs led to the use of the two techniques described below in two recent studies of small particles.

In one procedure, used in a study of colloidal particles, a regular collodion grating replica was prepared and was mounted on a specimen screen. A small droplet of the sol was then placed on the replica surface. After evaporation of volatile components, the specimen was ready for observation.

In the second method, used in the study of a dry, powder-like material, the particles were mechanically dispersed in an amyl acetate solution of collodion. This mixture was then used in place of the plain collodion solution in making a replica of a grating.

In both of these methods, the replica was prepared in a manner similar to the cellulose-taper-stripping technique described by Schaefer.[1]

In both cases the final image shows the particles with a grating as a background. In this way each micrograph carries its own calibration as seen in Figs. 1 and 2. Furthermore, electron-optical distortions are partially compensated by similar distortions in the grid-like background. However, it will be noted that, while each of these two methods affords a relatively direct means of determining magnification, each still remains subject to inaccuracies caused by such factors as physical distortion of the replica and non-perpendicularity between the electron beam and the replica.

[1] V. J. Schaefer, Science **97**, 188 (February 19, 1943).

New Books

Matter and Light. The New Physics.

By LOUIS DE BROGLIE. Pp. 300, 5×7¾ in. W. W. Norton and Company, New York, 1946. Price $2.75.

This is an American printing of the translation by W. H. Johnston of Matiére et Lumiére which was originally published in French in 1937. It duplicates the English translation which appeared in 1939.

It consists of a series of twenty-one brilliant essays dealing with modern physics and with philosophical studies raised by these new developments. They are grouped under six headings, as follows: A General Survey of Present-Day Physics, Matter and Electricity, Light and Radiation, Wave Mechanics, Philosophical Studies on Quantum Physics, and Philosophical Studies on Various Subjects. Only two of the essays contain mathematical equations which are, however, familiar to students of modern physics. Since the essays are quite independent of one another, the major portion of the book may be read by anyone who is interested in the revolutionary ideas arising from the development of quantum theory, relativity, and wave mechanics. The line of thought leading to the foundation of wave mechanics by the author is retold in several of the essays with interesting variations.

In this book, the scientist in other fields than physics, the philosopher, and the cultured amateur may find a delightful popularization of the principal ideas of the new physics. The physics student will be pleased by the masterly resumé of modern physics and by the author's insight into related philosophical problems. However, not the least profit to be gained by reading this book is in

regard to cultural values, to which the author is keenly sensitive. For example, the final essay, directed to science students, contains the following charge: "the younger generation must undoubtedly acquire learning so that they shall be able to benefit by past experience and by the stock of gathered knowledge; but beyond learning, they must acquire a love of beauty in all its forms, of the art of thinking, and of the expression of thought" and further—"without a zeal to pursue all that is loftiest in the intellectual, the esthetic, and the moral sphere—a civilization, however great the perfection of its material aspects, would soon be no better than a complicated form of barbarism." The intervention of a world war and the invention of the atomic bomb since the first printing of this book has only served to make it more worthy of careful study.

JOSEPH VALASEK
University of Minnesota

New Booklets

The Ohmite News for October 1946, published by Ohmite Manufacturing Company, 4835–41 Flournoy Street, Chicago 44, Illinois, features a brief biographical sketch of George Westinghouse.

Interchemical Review for summer 1946, published by Interchemical Corporation, 432 West 45 Street, New York 19, New York, has the following table of contents:

High-Speed Microtoming.....Mary C. Schuster
What a Research Library Offers.....Lisa G. Otto
Applying Pigment Colors to Textiles.....Winn W. Chase

The General Radio Experimenter, published by General Radio Company, 275 Massachusetts Avenue, Boston 39, Massachusetts, features in its October issue an article on "A Peak-Reading Voltmeter for the U-H-F Ranges."

Talk-A-Phone Company, 1512 South Pulaski Road, Chicago 23, Illinois, announced its new 1946 catalog, describing "the world's most advanced and complete line of inter-communication." 12 pages.

Leeds and Northrup Company, 4908 Stenton Avenue, Philadelphia 44, Pennsylvania, publishes a quarterly newspaper-style bulletin entitled *Modern Precision*. Its eight pages contain many short articles and photographs.

Fischer and Porter Company, Department 6F-7, Hatboro, Pennsylvania, has announced its Bulletin 40-A, which describes completely armored variable-area flow rate meters. 8 pages. Available for selected distribution to users of high pressure flow rate instruments.

Battelle Memorial Institute, Columbus, Ohio, recently published a brochure entitled *Research in Action*, a pictorial presentation of the work being done at Battelle in cooperation with industry. 58 pages, many photographs.

Engineers Specialties Division, 980 Ellicott Street, Buffalo 8, New York, has issued a four-page leaflet about its optical projection comparators. The folder illustrates a few examples of complete gauging.

Bakelite Review for July 1946, published by Bakelite Corporation, 30 East 42 Street, New York 17, New York, features an article on the use of expanded structural plastics in modern housing. 28 pages, many photographs.

Electro-Voice, Inc., Buchanan, Michigan (formerly South Bend, Indiana), has announced in its *Bulletin No. 131* the new cardyne cardioid dynamic microphone which it manufactures.

The Gaertner Scientific Corporation, 1201 Wrightwood Avenue, Chicago 14, Illinois, has issued its *Description M404*, dealing with the Gaertner M404 Ring Spherometer. 4 pages, free on request.

Dow Corning Corporation, Midland, Michigan, recently published an 8-page pamphlet entitled *DC 702 and DC 703, Silicone Fluids for Use in High Vacuum Diffusion Pumps*. This is a revision of the bulletin issued on April 15, 1946.

Farrand Optical Company, Bronx Boulevard and East 238 Street, New York 66, New York, has issued a four-page folder announcing its interference filters.

Plastics Divisions of General Electric Company, 1 Plastics Avenue, Pittsfield, Massachusetts, have issued a 14-page booklet entitled *What Are Plastics?* Numerous photographs. Written for the layman. Free on request.

General Radio Company, Boston 39, Massachusetts, recently issued a new 6-page bulletin, *New Variac Continuously Adjustable Transformers*.

Electro-Voice, Inc., 1239 South Bend Avenue, South Bend 24, Indiana, manufacturer of microphones and stands, has published a complete new catalog and selection guide. Free on request.

The Navy's Office of Naval Research, Washington 25, D. C., has republished tables computed by the Mathematical Tables Project of New York for the Massachusetts Institute of Technology Underwater Sound Laboratory of Division 6, National Defense Research Committee. The report is entitled *Scattering and Radiation from Circular Cylinders and Spheres—Tables of Amplitudes and Phase Angles*. The tables were originally computed for a limited audience. The Office of Naval Research believes the report will be of value to scientists engaged in acoustical and electromagnetic wave research, and will therefore make a limited number of these tables available without cost. Scientists wishing to obtain copies of the report should address the Office of Naval Research.

Contributed Original Research

Ion Beams in High Voltage Tubes Using Differential Pumping

E. S. LAMAR* AND W. W. BUECHNER
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received July 31, 1946)

Tests on a hydrogen ion gun assembly employing differential pumping are described. Focused beam currents as high as 215 microamperes were realized at the target end of a high voltage tube 6 feet long operating in the range from 300 to 400 kilovolts. Factors which limited the current to 215 microamperes are discussed and suggestions for realizing greater beam currents are presented.

INTRODUCTION

A NUMBER of investigations of the production and focusing of hydrogen ion beams have been made at the Massachusetts Institute of Technology,[1-5] the ultimate end in view being the use of ion beams together with high voltage vacuum tubes for investigations of the atomic nucleus. In references 4 and 5, which were published together, studies of ion sources and focusing systems were presented which led to a proposed ion gun design using differential pumping. This ion gun, shown in Fig. 1, was built and tested in conjunction with a high voltage vacuum tube. This paper is an account of the tests of the ion gun and vacuum tube assembly which were completed before the recent war interrupted the research, and the conclusions which were reached as a result of these tests. In the sections of this paper which follow, the accounts of the tests of the various parts of the assembly are presented in much the same order in which they were made.

THE ION SOURCE

The ion source was of the metal capillary type as described in reference 3. The capillary was 3.0 mm in diameter and 3.0 mm long. The outlet was 0.80 mm in diameter. In order to eliminate the need for water cooling, the arc chamber was made from a solid block of copper. The two cathode leads and the one for the anode were brought in through small spark plugs. Both cathode and anode ends of the arc chamber were closed by glass windows to facilitate visual examination of the arc.

The results of the total ion current measurements were in general agreement with those presented in reference 4. They are shown here in Fig. 2. The points shown are the averages obtained from a number of runs, the vertical lines giving the limits of the various measurements. On rare occasions, the output currents were outside the limits shown and approximated those characteristic of the glass ion sources described in reference 4. These rare runs have not been included in Fig. 2. The small fluctuations indicated in Fig. 2 and the larger ones just described are believed to result from changes in surface conditions as was discussed at some length in reference 4. A comparison of the ion currents obtained from the metal source with those to be expected from a glass one is sufficient in itself to show the desirability of completing the engineering development needed to produce a rugged glass ion source. This need is further sub-

* At present on leave at the Operations Evaluation Group, U. S. Navy.

[1] E. S. Lamar and O. Luhr, Phys. Rev. **46**, 87 (1934).

[2] Lamar, Samson, and Compton, Phys. Rev. **48**, 241 (1935).

[3] Lamar, Buechner, and Compton, Phys. Rev. **51**, 936 (1937).

[4] Lamar, Buechner, and Van de Graaff, J. App. Phys. **12**, 132 (1941).

[5] Buechner, Lamar, and Van de Graaff, J. App. Phys. **12**, 141 (1941).

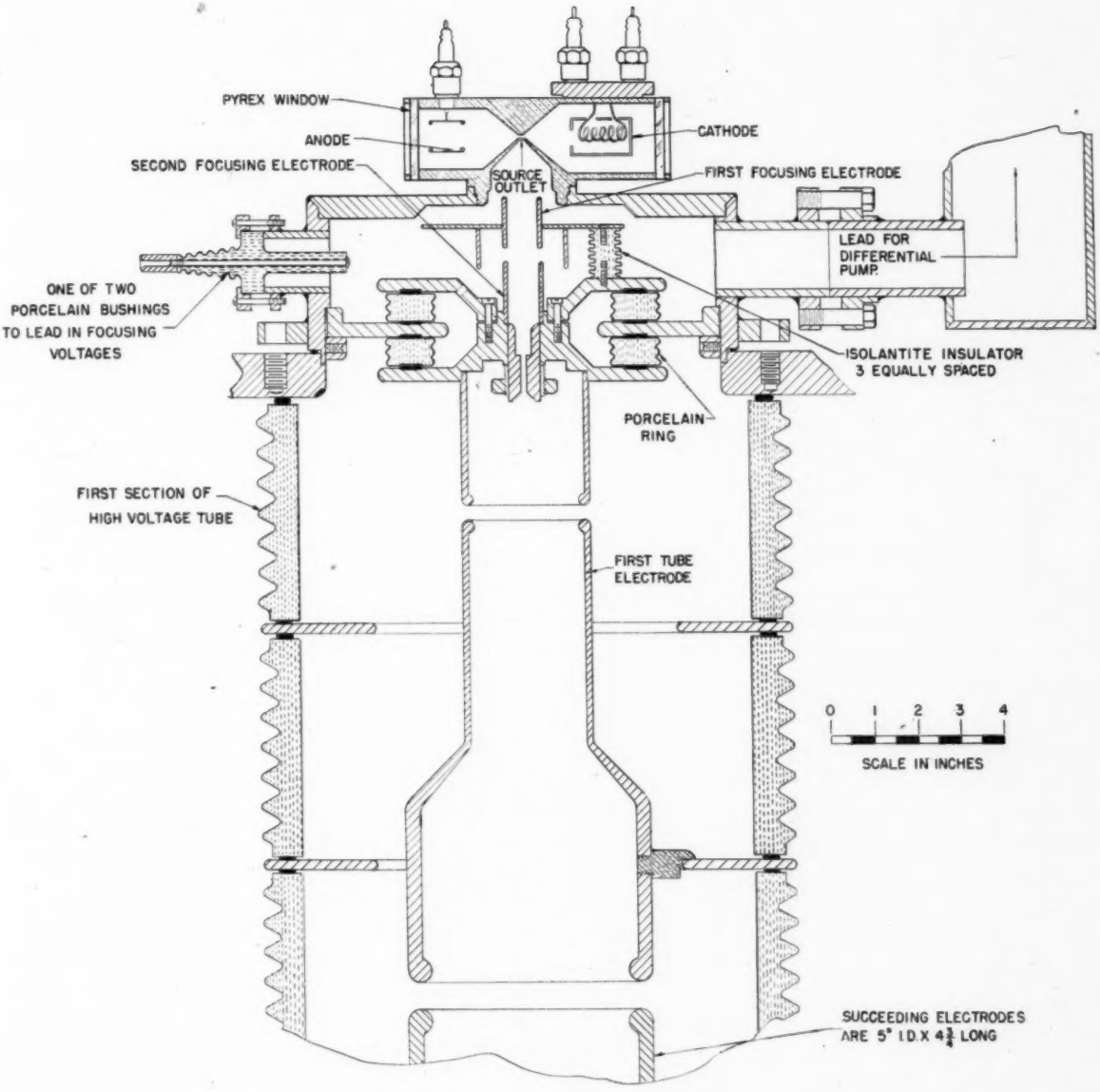


FIG. 1. Ion gun and initial focusing electrode of accelerating tube.

stantiated by comparing the atomic ion yields of the two types as shown by the mass analysis reported in reference 4.

THE INITIAL FOCUSING SYSTEM

The initial focusing system, shown in Fig. 1, had exactly the same electrode dimensions as the one described in reference 5. The outlet leading to the high voltage vacuum tube consisted of a canal $\frac{1}{8}$ inch in diameter and 1 inch long. In most of the experiments the accelerating potential was applied in two steps: between the focusing cone and first cylinder, and between the first and second cylinders; i.e., a double rather than a single lens was used. With this arrangement, for any given current, greater ion currents through the canal and into the high voltage tube could be realized than with the single lens system. As was discussed in reference 5, for a given total focusing voltage, the aberrations at the source outlet are less for a double lens than for a single one.

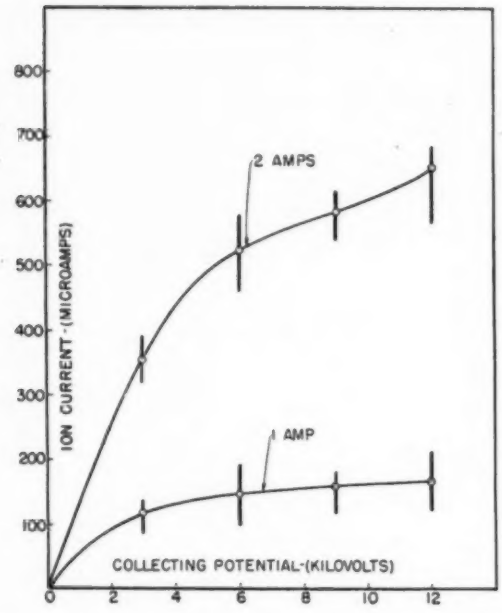


FIG. 2. Output ion current as a function of collecting potential for arc currents of 1 ampere and 2 amperes.

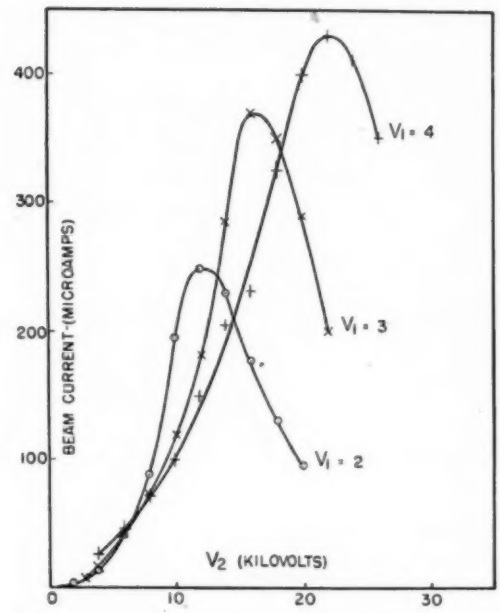


FIG. 3. Ion current focused through differential pumping canal into the high voltage tube as a function of the initial focusing potentials. Arc current = 2 amp.

Figure 3 shows a typical set of measurements of ion current entering the high voltage tube through the differential pumping canal. These measurements were made with all the tube electrodes connected together to form a deep Faraday collector. The potentials V_1 and V_2 of the first and second cylinders, respectively, were measured with respect to the focusing cone. It is to be noticed in Fig. 3 that for each value of V_1, the value of V_2 which gives the greatest beam current through the canal and into the high vacuum of the accelerating tube is quite critical. The ratio of voltages for best focus is a function of the first cylinder voltage V_1 and of the maximum output beam current. Since the maximum voltage of the power supply for the second cylinder was limited to about 26 kilovolts, it was not possible, in all runs, to read the point of best focus. For those runs in which the peak was reached, the ratio V_2/V_1 at the peak is given in Table I as a function of V_1 and the maximum beam current. As can be seen from the table, V_2/V_1 increases with increasing current and with decreasing voltage. In other words, the greater the space charge in the beam as it enters the second lens, the greater the second lens action required.

THE PUMPING SYSTEM

The vacuum pump attached to the critical focusing region was of the tapered nozzle type and that for the high vacuum tube was a steel pump of conventional design. Both these pumps used mercury and were provided with dry ice traps. An idea of the net pumping speeds of each of these pumps can be obtained from ratios of pressures in various regions. For the first pump, computing the pumping speed from the ratio of pressures between arc and initial focusing chamber involves a slight complication. The arc

TABLE I. V_2/V_1 for best focus.

Current (milliamps)	$V_1=2$ (kilovolts)	$V_1=3$ (kilovolts)	$V_1=4$ (kilovolts)	$V_1=6$ (kilovolts)
1.57			5.50	
0.75	6.75			
0.43			5.50	
0.37		5.50		
0.36	6.00			
0.33		5.33		
0.31			5.50	
0.25	6.00			
0.23				3.67
0.20			4.00	
0.19		5.00		
0.18			4.88	
0.16	5.25			
0.12	4.75			

pressures at which the measurements were made were such that the mean free path of the gas molecules in the arc was of the order of the outlet diameter. Hence the speed of the outlet was not constant but was a function of the arc pressure. It was found that the pressure ratio between initial focusing region and the arc was

$$R=5.03(10^{-3})+1.88(10^{-2})P_a, \qquad (1)$$

where R is the pressure ratio and P_a is the arc pressure. Using the low pressure part, i.e., $5.03(10^{-3})$, of this ratio, and the speed of the arc outlet, 5.0×10^{-2} liter per second for air, for low pressures, gives 10.0 liters per second for air as the speed of the intermediate pumping system.

The ratio of pressures between intermediate and high vacuum regions was 162. Since the speed of the $\frac{1}{8}$-inch canal was about 0.13 liter per second for air, that of the high vacuum pumping system was about 21 liters per second for air.

THE HIGH VOLTAGE TUBE AND GENERATOR

The high voltage tube was the one employed by Hill, Buechner, Clark, and Fisk.[6] For these tests the tube was mounted vertically with the source and pumps at the bottom at ground potential. The tube electrodes used by these authors were replaced by similar ones 5 inches in diameter except at the ion gun end. Here the first two electrodes were combined into one. The end of this electrode away from the ion gun was 5 inches in diameter. Several different electrode diameters were tried for the gap adjacent to the ion gun. The one finally adopted was 3″ as is shown in Fig. 1.

At the target end of the tube, several electrode arrangements were employed. The one finally adopted consisted of a deep "Faraday ice pail" having an insulated molybdenum plate at the far end and a two inch diameter opening facing the high voltage tube. A Pyrex window beyond the molybdenum plate made the plate visible from normal eye level. The molybdenum plate was connected to the end plate of the high voltage tube through a bank of B-batteries and a microammeter. The B-batteries provided a retarding potential of 135 volts to prevent secondary electrons from leaving the molybdenum plate.

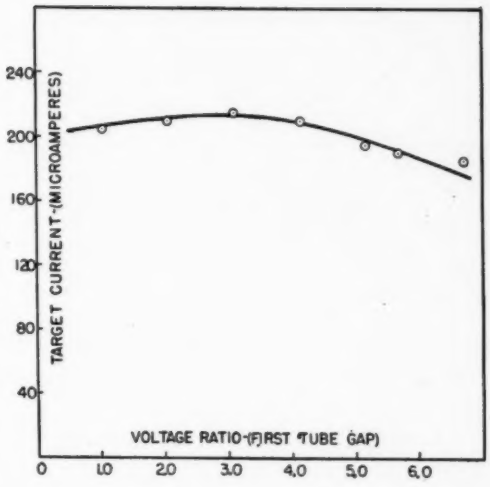


FIG. 4. Target current as a function of the ratio of the potentials across the first gap in the high voltage tube.

The large corona shield which covered the end plate and auxiliary apparatus was made partly of metal gauze to facilitate visual examination of the meter and of the molybdenum plate. The high voltage generator employed was of the belt type.[7] It was suspended from the ceiling and was connected to the corona shield by means of a length of stove pipe.

TESTS OF HIGH VOLTAGE TUBE AND ION GUN ASSEMBLY

In order to determine the effect of the first high voltage tube gap on the focusing of the ion beam, several different pairs of electrodes were tried. The smallest of these was $1\frac{3}{4}$ inches outside diameter. Corona control permitted independent adjustment of the potential of the first high voltage tube electrode. Variations of this potential over quite wide limits showed little or no effect on the ion current collected at the target end of the tube. In other words, the first high voltage gap adds little or nothing to the focusing of the ion beam. This is shown in Fig. 4 where the ion current collected at the target end is presented as a function of the ratio of potentials on the two electrodes of the first gap in the high voltage tube.

[6] Hill, Buechner, Clark, and Fisk, Phys. Rev. **55**, 463 (1939).

[7] We are indebted to Professor J. G. Trump of the Electrical Engineering Department for making this generator available to us for these tests.

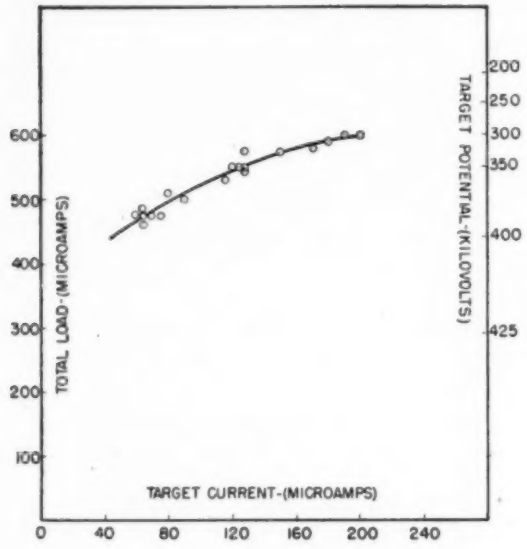


FIG. 5. Total tube current as a function of focused beam current.

The molybdenum plate was employed to give an indication of the state of focus of the ion beam when it reached the target end of the tube. With currents of the order of those presented in Fig. 4, a luminous spot, about $\frac{1}{4}$ inch in diameter, developed in the center of the plate. Because of radial heat conduction from the focal area, the luminous spot size serves only to set an upper limit to the diameter of the focal area. Since this was less than $\frac{1}{4}$ inch, the state of focus at the target end was considered satisfactory.

Attempts were made to realize currents at the target end of the tube greater than the 215-microampere maximum shown in Fig. 4. These attempts were not successful. Whenever the ion gun potentials or the arc current were increased beyond the values corresponding to this maximum, the generator potential dropped very rapidly, the high vacuum tube pressure increased, and electrical breakdown soon followed. These effects indicated the presence in the high voltage tube of a current in addition to that in the ion beam and the increase in pressure indicated that this parasitic current was, in some way, associated with the tube electrodes. The parasitic current may have its origin in the gas, at the electrodes, or both. In this connection, the results of Crenshaw[8] on the loss of energy of hydrogen ions in traversing various gases are of interest. If, under the present conditions, appreciable ionization takes place in the gas, certainly the positive products are not subsequently focused as is indicated by the fact that the target currents are about those to be expected from the primary beam. If the products of ionization strike the electrodes, they may be greatly multiplied by the emission of secondary electrons,[6] and large electron currents may thus originate from the electrodes. Such electron currents may be further augmented by photoelectrons resulting from excitation of the gas along the beam, since it has been observed, with ion beams in glass tubes, that the ion beam is quite luminous even at pressures as low as 10^{-5} mm of mercury. There is the further possibility that secondary electrons may be liberated from the electrodes by primary ions scattered from the walls of the canal. This possibility was ruled out in the following way: A new canal, consisting of a diaphragm facing the source and a slightly larger diameter canal facing the high voltage tube, was installed. Although this new canal possessed a much reduced scattering area, the results obtained using it were the same as those with the previous arrangement.

An attempt was made to measure the total current in the high voltage tube, as a means of evaluating the parasitic current. Since this measurement involved certain assumptions, it can be considered as indicative only. First of all, it was assumed that a one to one correspondence existed between generator voltage and generator load. A calibration of the generator was then run using a corona load which could be measured. With this calibration, a measurement of generator voltage provided a measurement of total current. In Fig. 5, this total current is presented as a function of beam current reaching the target. The generator voltage, corresponding to each total load current, is given on the right-hand edge of the diagram. It can be seen from Fig. 5 that, subject to the original assumption, the total load exceeds the beam load by some 400 microamperes.

It is clear that if beam currents of more than a few hundred microamperes are desired, the parasitic current must be reduced. Since the primary cause of this current is collision, either

[8] C. M. Crenshaw, Phys. Rev. **62**, 54 (1942).

scattering, exciting, or ionizing, in the gas, it is essential that the gas pressure in the high voltage tube be kept as low as possible. If the two pumps of 10 and 20 liters per second, respectively, here employed had been replaced by a single pump used in the conventional manner, a net pumping speed at the source of approximately 1600 liters per second for air would have been required to maintain the same hydrogen pressure in the accelerating tube. Any multiplication of parasitic current by secondary electron emission from the electrodes can be reduced by introducing a diaphragm in the center of each electrode. Since the center is an equipotential plane, no distortion of field should result and the secondary electrons formed off the axis would be stopped after a maximum travel of one tube gap. The focusing results indicate that the hole in the center of the diaphragm could be as small as $\frac{1}{4}$ inch in diameter without cutting off the ion beam. If this were done, it would of course be necessary to arrange the rest of the tube geometry so as to maintain an adequate pumping speed, particularly if the pumping were done from the target end of the tube. Time did not permit investigation of this possibility in the present research. It is, however, strongly recommended for further work.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor R. J. Van de Graaff for his continued interest in the ion source work and for much helpful discussion. We wish also to express our appreciation to the other members of the high voltage group at the Institute for their suggestions and help in many ways. We are glad to have this opportunity to express our gratitude to the Research Corporation and to the Carnegie Corporation of New York for their generous grants supporting this work.

X-Ray Diffraction Studies of Color Variation of Iron Oxide Pigments

H. Birnbaum, H. Cohen, and S. S. Sidhu

Cooperative X-Ray Laboratory, University of Pittsburgh, Pittsburgh, Pennsylvania

(Received August 16, 1946)

Samples of synthetic iron oxide (Goethite) taken at regular intervals from a twelve-day continuous process showed progressive change of color, ranging from pale yellow to dark brown. X-ray diffraction, electron microscope, and optical microscope studies lead to the conclusion that the color change in these samples is mostly caused by particle growth.

THE process of producing synthetic iron oxide pigments has been reviewed by Mattiello.[1] A characteristic feature of these pigments is the progressive change of color from

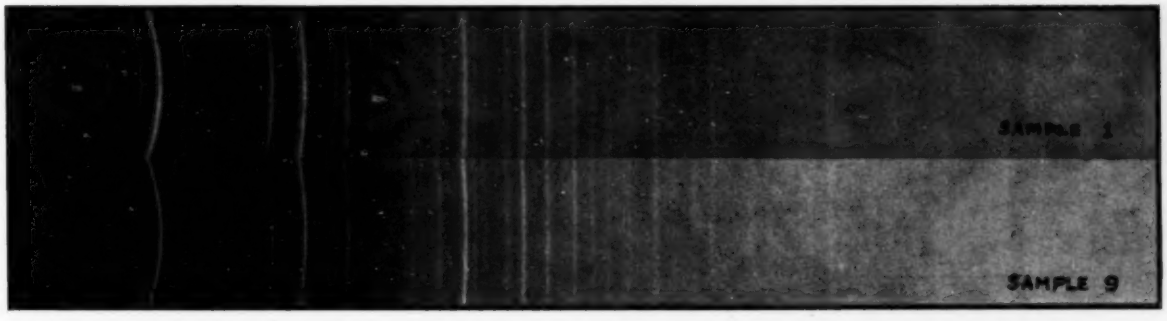


Fig. 1. X-ray diffraction patterns of samples 1 and 9. FeK unfiltered radiation, $\lambda = 1.933$A. Diameter of camera 143.2 mm.

[1] J. J. Mattiello, *Protective and Decorative Coatings* (John Wiley and Sons, Inc., New York, 1942), Vol. II, p. 300.

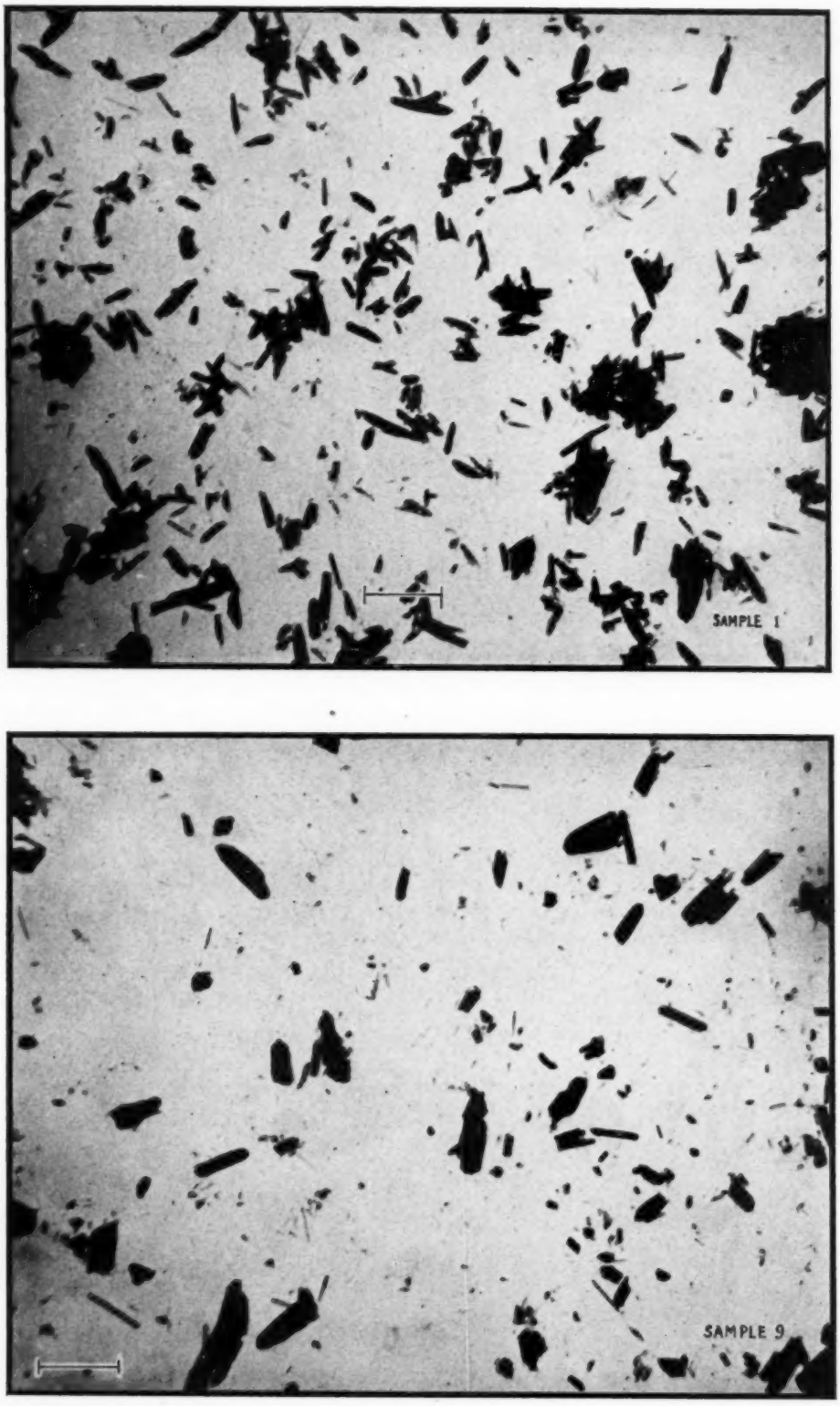


FIG. 2. Electron photomicrographs of samples 1 and 9. Magnification of sample 1: 11250×, and of sample 9, 12000×.

pale yellow to dark brown, as may be observed during processing. The commercially workable methods employ ferrous sulfate, scrap iron, air and steam. According to Fireman[2] a small amount of ferrous sulfate, about 1 percent to 2 percent, occasionally up to 3 percent, is converted into ferric sulfate.

"This suggests that the ferrous hydroxide from the partly hydrolyzed ferrous salt is first attacked by the oxygen of the air, causing the separation of the hydrated ferric oxide and releasing a corresponding amount of acid, which goes to form a certain amount of ferric salt. The latter is then reduced by the iron, which in dissolving restores the ferrous salt. Thus, the new incoming air finds the conditions of the solution exactly the same as the air which preceded it and repeats the same oxidizing action."

Since the chemical analysis of different colored pigments gave approximately the same average composition of 88.0 percent Fe_2O_3 and 12 percent H_2O, x-ray diffraction and electron microscope studies were undertaken to determine the nature of color variation. The pigments used in these studies were prepared as follows.

EXPERIMENTAL

To 1100 cc of 9 percent $FeSO_4$ solution were added 200 g NaOH (89.4 percent purity) dissolved in 1000 cc water, under stirring at room temperature. The total volume was brought up to 10 liters and the gelatinous ferrous hydroxide oxidized by slow agitation. The crystallites of ferric oxide monohydrate thus obtained, including mother liquor, were transferred to a tank containing 70 liters of 5.7 percent ferrous sulfate solution and 22 kilograms of scrap iron. As the scrap iron was progressively converted into pigment, more metal was added to maintain a continous process.

The temperature in the reaction tank was kept between 60–65°C during the twelve days of continuous operation. The oxygen, necessary for the reaction, was supplied by forcing considerable air through the solution. Samples, given in Table I, were taken at stated intervals, then filtered, washed, and dried to constant weight at temperature of 125°C. X-ray diffraction patterns of the samples were made by the usual Debye-Scherrer-Hull method employing unfiltered Fe*K* radiation. In addition, all samples were examined with optical microscope for particle size, and electron photomicrographs were made of samples 1 and 9.

TABLE I. Color of pigments, depending on reaction time.

No. of sample	Color	Reaction time
1	Light lemon yellow	24 hours
2	Lemon yellow	48
6	Orange yellow	168
8	Light reddish brown	216
9	Reddish brown	268

DISCUSSION

X-ray diffraction patterns of samples in Table I were identified as those of the mineral Goethite.[3] Comparison of these patterns showed that the deepening of color in the pigment was accompanied by an increasing sharpness in the diffraction lines. This is illustrated by the photographs of samples one and nine in Fig. 1. Examination of the samples with the optical microscope disclosed a distinct difference in particle size. This was subsequently confirmed by the electron photomicrographs as shown in Fig. 2. While increase of sharpness and greater resolution in x-ray diffraction lines are, in general, an indication of grain growth, or strain free material or both, it appears from x-ray diffraction patterns, electron photomicrographs, and optical microscope studies that the progressive color change of the iron oxide pigments studied here is a visible expression mostly of the particle growth.

ACKNOWLEDGMENT

The authors want to thank H. C. O'Brien, of the University of Pittsburgh, for the preparation of the electron photomicrographs. The contributions of M. J. Pooley of Hachmeister-Inc., Pittsburgh, Pennsylvania, who assisted in preparing the pigment samples, are also acknowledged.

[2] P. Fireman, Ind. Eng. Chem. **18**, 286–287 (1926).

[3] C. W. Bunn, J. Sci. Inst. **18**, 70–74 (1941).

Mass Spectrometer for Leak Detection

A. O. NIER,* C. M. STEVENS,** A. HUSTRULID,*** AND T. A. ABBOTT†
The Kellex Corporation, New York, New York
(Received June 27, 1946)

A simple low resolution mass spectrometer is described which has been used successfully in detecting small leaks in high vacuum equipment. One part of He in 200,000 parts of air can be detected.

1. INTRODUCTION

THE problem of finding small leaks in a high vacuum system constructed of metal or glass and metal has always been a difficult and tedious task. Many methods for facilitating the location of leaks have been studied and described.[1,2] Most of the methods depend upon looking for a change of pressure in the system when the part suspected of leaking is waxed or closed with some other substance; or by looking for a change in the nature of the gas or vapor in the system when the part under test is painted with ether, CCl_4 and other liquids or when a stream of gas, such as CO_2, H_2, He, is directed on it. The device for indicating changes in pressure or in the nature of the vapor mixture must be very sensitive in order to find small leaks in systems having many leaks or in systems having surfaces which outgas and hence act as virtual leaks.

Of the various methods for hunting leaks, the scheme of directing a stream of relatively inert test gas at the suspected part of the apparatus and then looking for changes in the concentration or the presence of this particular gas in the system under test has many advantages. Some of these are: (1) gas will not contaminate the system or the pumps; (2) it will enter the system readily through small leaks; (3) there is no danger of closing a leak and having it open some time later; and (4) once a leak is found, there is no time lost in continuing the hunt because the test gas entering the system is quickly pumped out. The ideal "leak detector" is then an instrument which continually samples the gas in the system and instantly notes the presence or change in concentration of the test gas regardless of the presence of other gases or vapors.

A method which has been used with success in the construction of large vacuum equipment employs improved vacuum techniques and a new leak-detecting instrument. The outstanding contributions have been the development of a special mass spectrometer and a method of using it with a tracer gas in the location of minute leaks.[3]

A typical leak-testing arrangement employing the tracer gas method and using a mass spectrometer as a detector is shown in Fig. 1. The system which is being tested is evacuated by the continuous action of the vacuum pumps. An atmosphere of tracer gas may completely surround the system or a small jet of it may be used to probe the suspected zones. Any tracer gas which leaks into the system is rapidly detected by the mass spectrometer which continuously samples the gas being pumped from the system under test. Helium is one of the best tracer gases because it has a low molecular weight, does not occur to any extent in air or in other gases, is not adsorbed on the surfaces, can be readily pumped out of the test equipment, and is easily procured.

2. THE MASS SPECTROMETER AS A DETECTOR

The first portable mass spectrometer designed for detecting leaks was built at the University of Minnesota in the spring of 1943. A more recent model was developed in The Kellex Corporation

* Present address: University of Minnesota, Minneapolis, Minnesota.
** Present address: Distillation Products Corporation, Rochester, New York.
*** Present address: University of Minnesota, Minneapolis, Minnesota. Dr. Hustrulid worked with A. O. Nier at the University of Minnesota on the first sample instruments which were built under OSRD Contract OEMsr-149.
† Present address: Standard Oil Company (Indiana), Chicago, Illinois.
[1] J. Strong, *Procedures in Experimental Physics* (Prentice-Hall, Inc., New York, 1942), pp. 134–7.
[2] E. J. Lawton, Rev. Sci. Inst. **11**, 134 (1940).
[3] R. B. Jacobs and H. F. Zuhr, J. App. Phys. **18**, 34 (1947).

laboratories and differs from the first model by including an all-metal spectrometer tube and diffusion pump and a special electrode at the collector end of the tube for decreasing the background ion current.

The several components of the mass spectrometer are shown in Fig. 1. The details of the all-metal spectrometer tube are shown in Fig. 2. The gas enters the tube at the top and is ionized by a stream of electrons which are emitted from the heated filament and accelerated toward the trap. The number of ions produced depends on the electron emission and energy and on the gas pressure in the ionizing electron beam.

The ions are pulled out of the electron beam and accelerated by an electric field between the shield and plate containing the source slit. The emission regulator contains the necessary electronic rectifying and regulating circuits for maintaining the desired potentials on the accelerating and focusing plates, as well as holding the electron emission constant at any desired value.

A magnetic field is produced between the iron pole pieces by a permanent magnet. Since the magnetic field is wedge-shaped, it has a refocusing property for divergent ion beams as well as a resolving property.[4–6] The magnetic strength is such that an ion accelerating voltage of about 300 volts is required for collecting helium ions.

After leaving the magnetic field the selected ion beam passes through two baffle plates and a suppressor plate before striking the collector plate. The intensity of the selected ion beam is measured by a conventional battery-powered, negative feedback amplifier with a milliameter as an output meter. As shown in Fig. 2, the first stage of the amplifier which consists of a 954 tube and a 10^{11}-ohm input resistor is mounted inside the spectrometer tube.

The pressure in the spectrometer tube is measured by an ion gauge whose electron emission is held constant by electronic regulating circuits. The ion gauge is provided with a special filament cut-off circuit consisting of an amplifier tube and a relay. This circuit protects the ion gauge and spectrometer tube filaments from burning out by automatically cutting off their currents when the pressure exceeds a predetermined value.

The tube is evacuated by a high speed oil diffusion pump. A liquid nitrogen trap between the pump and spectrometer tube is used to achieve a residual pressure of 2×10^{-6} mm Hg in the tube. The pumping speed measured at the ionization gauge is approximately 30 liters/sec. This insures a rapid response of the instrument to changes in tracer gas concentration at the throttling valve. In probing, an indication is obtained on the output meter within a few seconds after the probe passes a leak.

3. PERFORMANCE OF THE LEAK DETECTING MASS SPECTROMETER

Although the instrument can be used to detect any one of several different gases by the adjustment of the ion accelerating voltage, it is normally tuned to indicate only He^+ ions. The selectivity of the instrument is shown by the mass spectrum of air containing a small amount of helium which is plotted in Fig. 3. The amplifier output current is plotted for the different values of the accelerating voltage. The heights of the current

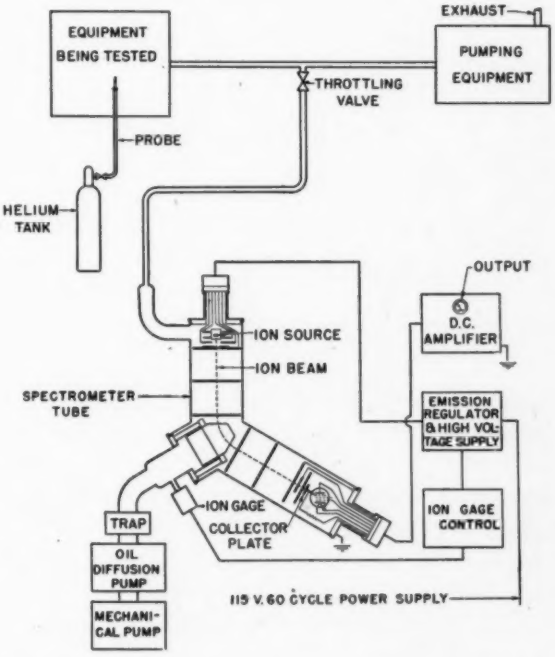


FIG. 1. Schematic diagram of typical leak testing set-up showing interconnections of mass spectrometer components and equipment to be tested.

[4] A. O. Nier, Rev. Sci. Inst. **11**, 212 (1940).
[5] W. E. Stephens and A. L. Hughes, Phys. Rev. **45**, 123 and 513 (1934).
[6] N. F. Barber, Proc. Leeds Phil. Lit. Soc. **2**, 427 (1933).

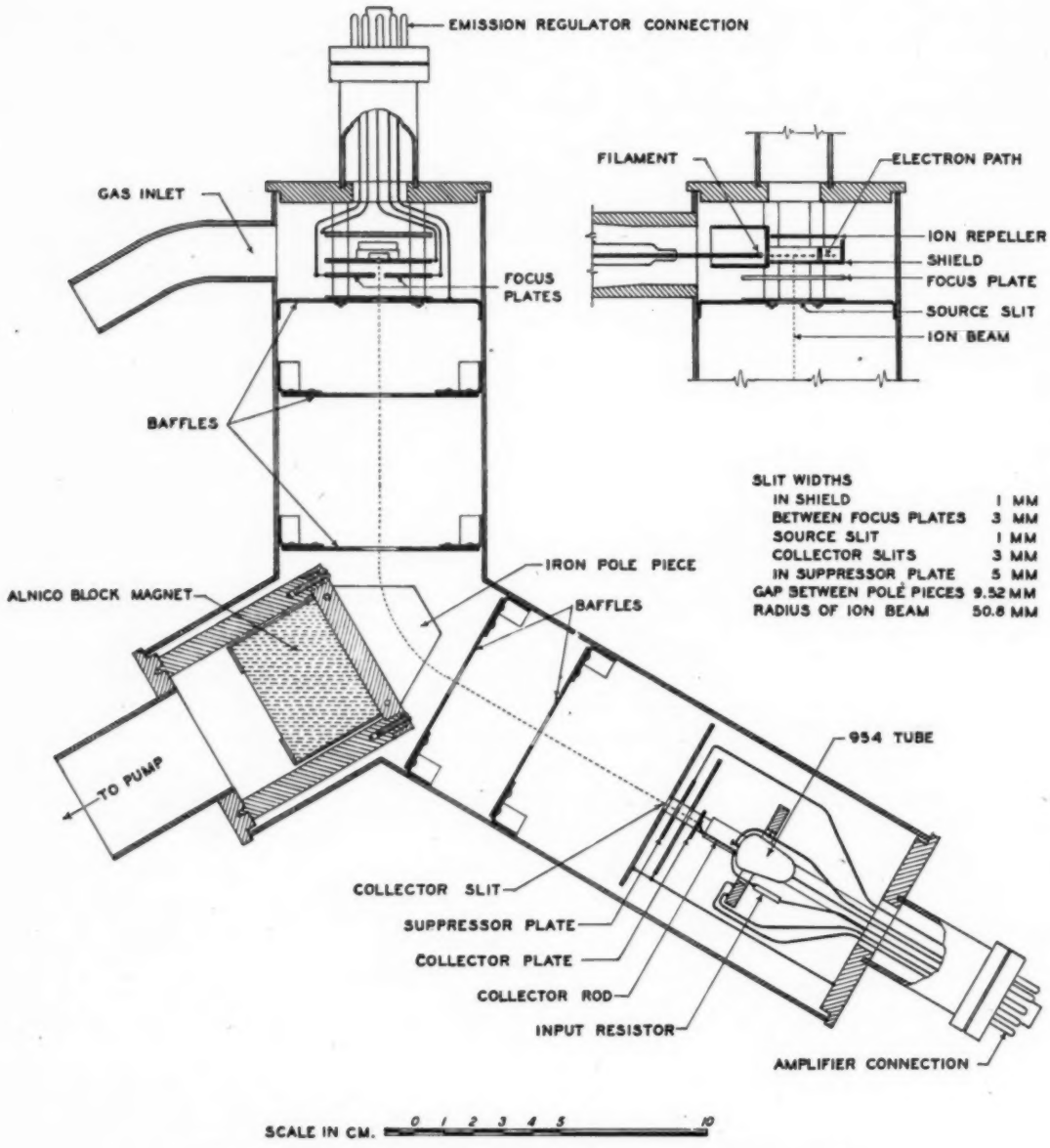


FIG. 2. All-metal spectrometer tube for leak detecting mass spectrometer.

peaks are proportional to the relative concentrations of the ionized gases in the sample. High sensitivity is made possible by two special features of the tube which will be explained below.

If the pressure in the tube is increased, the peaks increase in height since more ions are produced. Unfortunately, increased pressure also greatly increases the width of the peaks at the base especially on the high voltage side of the peak. This produces an increase in background which more than offsets the advantage gained because of the increased height of the peaks. The background is increased because at the higher pressure more nitrogen and oxygen ions lose energy by collision in passing through the region between the source and the magnet and are collected at the same time as are ions of lower mass.

The insertion of a baffle between the ionization chamber and the rest of the tube insures a high pressure in the ion source without a correspond-

ingly high pressure in the rest of the tube. Thus, this baffle materially increases the effective sensitivity of the tube.

A further increase in sensitivity has been obtained by the use of a suppressor plate which is placed just ahead of the collector plate. This plate is operated at a potential near that of the ionizing region so that those ions which have lost energy by collision cannot pass through the slit in the suppressor plate and are not collected on the collector plate. Thus, oxygen and nitrogen ions which have lost energy and would otherwise be collected along with He^+ are repelled, and the background is again decreased to materially increase the sensitivity to helium as shown in Fig. 4. Moreover, the He^+ ions in passing through the suppressor plate lose almost all their energy so that they produce only a negligible number of undesirable secondary electrons.

Under normal conditions the throttling valve connecting the spectrometer was opened until the ion gauge indicated a pressure of 10^{-5} mm in the spectrometer pump lead. This corresponded to an ion source pressure of 2×10^{-4} mm. With an electron emission of 5 ma and a He to air mixture of 1 to 10,000 a He^+ ion current of 6×10^{-12} amp. was obtained. This corresponded to a full scale deflection of the output meter. Since it was possible to detect conveniently He^+ peaks as small as 5 percent of full scale reading of the output meter, the sensitivity of the instrument was said to be 1 in 200,000.

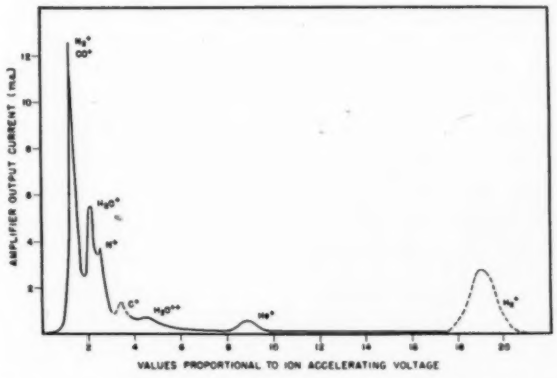


FIG. 3. A typical mass spectrum of air and helium mixture taken on a leak detecting mass spectrometer.

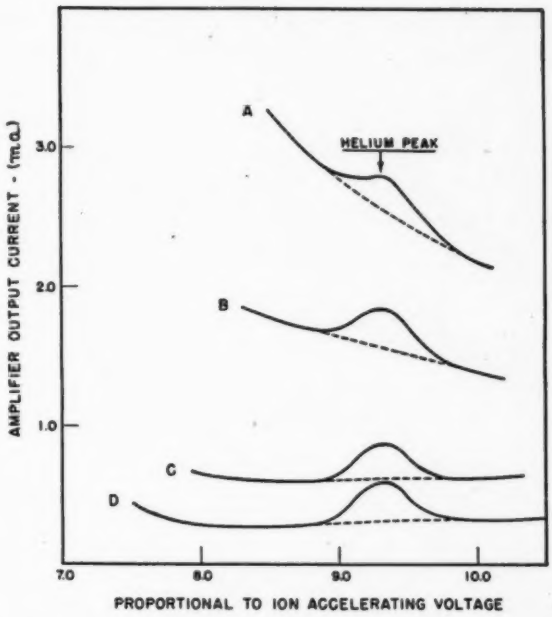


FIG. 4. Effect of suppressor field on mass spectrometer background current in the region of the He^+ peak for a sample having one part He to 50,000 parts of air. Pressure in spectrometer tube body $= 1.1\times10^{-5}$ mm Hg; pressure at ion source $= 20\times10^{-5}$ mm Hg.

Curve	Volts
A	0
B	+175
C	+250
D	+400

Energy of helium ions approximately 330 volts. Helium ions are collected in case of curve *D* in spite of high suppression voltage due to fact that slit in suppression plate is wide and hence potential of space in slit opening is lower than 400 volts because of proximity of grounded plates on either side.

In order to reduce the sensitivity of the apparatus it was possible to reduce the electron emission or add resistance to the output meter circuit. A sensitivity reduction of a factor of 1000 could be obtained.

4. ACKNOWLEDGMENTS

The authors wish to acknowledge the many contributions made by R. B. Thorness to the design and development of the leak detecting spectrometer. This paper is based on work done for the Manhattan Project under Contract No. W-7405 Eng. 23 at the Kellex Corporation and on earlier work at the University of Minnesota under OSRD Contract OEMsr-149.

New Developments in Vacuum Engineering†

ROBERT B. JACOBS* AND HERBERT F. ZUHR**
The Kellex Corporation, New York, New York

(Received June 12, 1946)

Methods developed for obtaining vacuum tightness in the K-25 Gaseous Diffusion Plant for the separation of U^{235} are described. A general account of the vacuum engineering of the K-25 Plant also is given. The dynamics of vacuum systems are described, including the use of the mass spectrometer leak detector on rapidly responding systems. New techniques described include, among others, the helium hood method for measuring vacuum tightness and the use of calibrated leaks.

1. INTRODUCTION

1.1 Degree of Tightness Required

THE development of leak detection techniques for the Gaseous Diffusion Plant for the separation of U^{235} was one of the major engineering problems connected with the project. A greater degree of tightness was required than had been achieved, hitherto, in any commercial plant, and the construction schedule did not permit the use of the time-consuming methods generally employed. It was recognized, accordingly, that vastly improved techniques were necessary to greatly increase the sensitivity and speed of testing, otherwise a prolonged period of testing would intervene between the completion of construction and the initiation of operations. After all known methods and several new ones were considered, the mass spectrometer was selected as the basic testing instrument. This paper describes the techniques for leak detection developed with the mass spectrometer, and indicates the general utilization of those techniques.

(*a*). *Process Specifications*

Process design required that the inleakage of ambient air into the process stream be no greater than one one-hundredth of a standard cubic foot per hour per thousand cubic feet of plant volume. This is equivalent to a pressure build-up at high vacuum of eight microns (0.00015 lb./sq. in.) per hour. Ambient air must be kept from the process stream, principally because of its water vapor content. Water causes destruction of the uranium hexafluoride, the products of the reaction plugging the minute pores of the barrier. To assure continued, efficient operation of the plant by reducing the effects of inleakage to a minimum, a large portion of the plant equipment was surrounded by air with a dewpoint of −40°F. Elaborate precautions still were necessary to assure that the tightness of the plant was maintained.

Equivalent tightness was required also in the heat exchange equipment, which might otherwise leak coolant into the process system.

(*b*). *Precedents*

Industrial precedent for vacuum tightness of the required degree had been limited to compact equipment of relatively small volumes, such as radio tubes, and some of the smaller mercury rectifiers. Indeed, the general trend in industry had been toward accepting leakage in vacuum systems and evacuating continuously with large, high speed diffusion pumps, in order to maintain sufficiently low pressures. The continuously pumped mercury arc rectifier, with a volume of about seventy cubic feet, composed of some fifteen separate parts, had constituted the largest piece of high vacuum equipment generally used on a commercial scale. A few mercury turbine installations had been made with inleakage approximately ten times that permitted the Gaseous Diffusion Plant. Perhaps the modern vacuum still for petroleum oil most closely approximated the complexity and size of the proposed plant, and these units had been considered tight, when the inleakage was approximately 500 times those

† This paper is based on work done for the Manhattan Project under Contract No. W-7405 Eng. 23 at the Kellex Corporation.

* Present Address: Distillation Products, Inc., Rochester, New York.
** Present Address: Hydrocarbon Research, Inc., New York, New York.

permitted by process specifications for the Gaseous Diffusion Plant.

1.2. Schedule *vs.* Effort Required

Early experience in the laboratories and pilot plants, working in the diffusion process, both here and in England, indicated that the existent testing methods, if applied to a full scale plant would require a testing effort (measured in man-hours) at least equivalent to the effort required to manufacture and assemble such a plant. Indeed, in many laboratories it was common practice, after several men of graduate caliber had devoted a month or two in an unsuccessful attempt to make an experimental set-up tight, to coat vessels with organic "vacuum" waxes. The requirement that the process stream be free from organic contaminants precluded the use of such subterfuge in the Gaseous Diffusion Plant. Further, the exigencies of the project would not permit such an uncertain method of excluding air from the process stream.

Through application of the leak detection techniques developed, using the mass spectrometer as the basic testing instrument, it was possible to improve the process specifications fourfold. A pressure build-up at two microns (0.00004 lb./sq. in.) per hour in the assembled plant was achieved at an effort measuring approximately one percent of the total effort required to manufacture and assemble the process plant.

1.3. Pressure Units

Pressure units employed in high vacuum work differ from those most commonly encountered in engineering practices. The unit most frequently used is the micron which is 0.001 mm of mercury absolute or 1/760,000 of standard atmospheric pressure (14.7 lb./sq. in. absolute). Hence, a micron corresponds to a pressure of 1.94×10^{-5} lb./sq. in. absolute. The so-called high vacuum range usually refers to that where the pressure is less than one micron.

1.4. Leakage Units

Leakage units are usually microns per hour, when referred to a particular vessel, or more generally, micron cubic feet per hour (mcfh). The first unit only has meaning when referred to a particular vessel. Multiplying the rate of rise of pressure in microns per hour due to inleakage by the volume of the vessel in cubic feet gives the value of the inleakage in micron cubic feet per hour. The latter unit completely defines the leak, since it is independent of the volume of the vessel, and from it the mass rate of inleakage may be computed directly. Hence, a leak of one micron cubic foot per hour means that the leak is of such value as to cause a rate of rise of pressure of one micron per hour in volume of one cubic foot.

2. KNOWN TESTING TECHNOLOGY AT THE BEGINNING OF THE PROGRAM

2.1. Pressure Methods

It was general practice to test process equipment operating under vacuum, using an internal test medium under pressure, in conjunction with an external indicator for location of the leaks. After all possible leaks were discovered and repaired using this procedure, the system was evacuated and the rate of pressure rise observed as a measure of the inleakage. If the pressure rise exceeded specifications, the system was repressured and the entire cycle repeated. These pressure testing techniques are discussed in more detail in Appendix 7.2. Where applicable to process service lines, these techniques were used in the construction of the Gaseous Diffusion Plant. Design of the process equipment would not permit the application of more than five lb./sq. in. gauge to most of the plant, further decreasing the applicability of pressure testing techniques. However, an improvement of several hundredfold would be necessary for use of this method in the testing program for the plant proper.

2.2. Vacuum Methods

Small pieces of vacuum equipment were usually tested using vacuum techniques. Such methods consisted of observing the change in apparent pressure in the system, as measured with either a hot wire or an ionization gauge when the nature of the residual gas in the system was changed by the substitution of a probe gas (or vapor) for the air normally entering through the leak. Most commonly used was the technique of observing a hot wire gauge, while the exterior of the vessel was sprayed with acetone. A more detailed ac-

TABLE I. Analysis of basic leak detection instruments.

	Mass spectrometer	Differential pirani gauge	Single pirani gauge	Pressure testing
Inherent instrument sensitivity (ratio of probe gas to air detectable)	$\frac{1}{100{,}000}$	$\frac{1}{3000}$	$\frac{1}{300}$	Bubble growth at 10 mm^3/hr
Operating pressure	<0.1 microns	<25 microns	<50 microns	45 psig
Speed of probing	5 ft./min.	As fast as area can be wet with spray gun	As fast as area can be wet with spray gun	Observation of smooth soap film for at least 5 min.
Smallest leak detectable on vessels having a pressure build-up at 200 microns per hour and a volume of:				
<1 cubic foot	<0.01 mcfh	0.10 mcfh	1.0 mcfh	5 mcfh
1–10 cubic feet	0.01–0.02	0.10– 1.0	1.0– 10.0	5
10–100 cubic feet	0.02–0.2	1.0 –10.0	10.0– 50	5
100–1000 cubic feet	0.2 –2.0	10.0 –50	50 –500	5

count of these methods is given in Appendix 7.3. Results obtained using these methods indicated that some improvement over pressure testing techniques could be reasonably expected. However, a corresponding increase in the caliber of the necessary personnel using these methods was also indicated, and the increase in the speed of testing was uncertain.

3. DEVELOPMENT OF DYNAMIC VACUUM TESTING TECHNIQUES

Study of the evidence on hand indicated that further advances in pressure testing methods would, at best, advance their useful sensitivity to equivalence with the then existent vacuum techniques, at the expense of an increase in testing time. It was evident that improvement of the existing vacuum techniques could have been effected, but again only at the expense of additional testing time. The increased time factor made both proposals most unattractive. However, new vacuum techniques were visualized which would permit a tenfold improvement in sensitivity and a simultaneous improvement in the speed of testing. These new techniques included:

1. Use of selective instruments; that is, instruments which give a nearly null reading for air and residual gases, responding only to a probe gas.
2. Use of these instruments dynamically; that is, as adjuncts to a high speed evacuating system permitting their use under optimum conditions.
3. Use of the selectivity of the indicating instruments to permit estimation of the amount of leakage even in the presence of outgassing and unrepaired leaks.

It was therefore decided to investigate and develop, if practical, a selective, continuously sampling leak detector, and to determine the optimum conditions for its use.

3.1. Leak Detectors Developed and Investigated

The Kellex Vacuum Engineering Group developed the Differential Pirani Gauge as a new continuous sampling unit, and investigated the use of the optical spectrometer as a continuous sampling unit. The University of Minnesota group had developed a compact, specialized mass spectrometer.[1] The use of the instrument, as a continuous sampling (or dynamic) instrument, was developed by the Kellex group. All three of these instruments were selective. Comparison of the Differential Pirani Gauge and mass spectrometer to older methods are made in Table I.

(*a*). *Optical Spectrometer*

The optical spectrometer, off-hand, would appear to be a suitable instrument for leak detection work. It is simple, its response is instantaneous, and its sensitivity for certain probe gases is very high. However, under actual working conditions, it is found to be unreliable due to the uncertain "clean-up" of the discharge tube used as a source after exposure to different gases.

(*b*). *Differential Pirani Gauge*

This instrument consists of two hot wire vacuum gauges known as Pirani Gauges. These

[1] A. O. Nier, C. M. Stevens, A. Hustralid, and T. A. Abbott, "Mass spectrometer for leak detection," Contribution from the Kellex Corporation, Research Paper C3.

are connected electrically in opposite arms of a Wheatstone bridge. The two gauges are connected to the vacuum line of the system under test by a short "tee" connection. A cold trap is placed on one of the arms of the tee leading to the gauges, but not on the other.

When air is the only gas present, the pressures in the two gauges are equal and the galvanometer reads zero. However, when a leak is probed with a liquid such as methanol, the latter reaches one gauge only, due to the presence of the trap in front of the other, and the Wheatstone bridge is thrown off balance. The instrument thus acts as a detector for condensible probe gases, being little affected by the presence of other gases.

(*c*). *Mass Spectrometer*

The mass spectrometer has been used by physicists for many years, principally to determine the isotopic constitution of different substances. In its usual form it is a large and rather complex instrument requiring a highly skilled operator.

A simplified instrument specifically designed for vacuum testing was developed at the University of Minnesota. A considerable reduction in size and weight was made, making a portable instrument possible. Several electronic controls were incorporated reducing the number of manual adjustments required. However, carefully trained operating and maintenance crews are required to use these instruments successfully for leak detection on a large scale.

In popular literature, the mass spectrometer is referred to as an "Atom-Sifter" due to the ability of the instrument to "sift out" and measure the concentration of atoms of any kind.

This characteristic of the mass spectrometer was the basis in choosing a probe medium. This medium should have the following properties:

(1) It should be unique in the mass spectrum of residual gas in the system under test and practically nonexistent in the normal atmosphere surrounding the equipment under test.
(2) It should be readily removed from the system by pumping and should not contaminate the system.
(3) It should be of low molecular weight and low viscosity, and should be readily available.

Helium fulfills all of the requirements and was chosen as the probe gas for use in the testing program.

The instrument works in the following manner: Referring to Fig. 1, the gaseous mixture to be analyzed is pumped into that part of the spectrometer which includes the source. When used as a leak detector, the spectrometer is fed through a throttle valve connected to the main vacuum lines. The pressure in the spectrometer is kept less than 10^{-4} mm by continuous pumping. In the source of the spectrometer, molecules of the different elements present are ionized by an electron beam. The number of ions of any particular element so produced in a function of the concentration of that element in the gaseous mixture fed into the spectrometer.

The ions thus formed are collimated and given a certain momentum toward the analyzer by suitable electrical potentials. On passing through the wedge-shaped magnetic field, the ions are deflected through certain angles which depend upon their respective masses.

In Fig. 1, the electrical fields are adjusted so that He^+ ions enter the collector where they are measured. Ions of greater or of smaller mass fall on either side of the collector slit and are not measured. Ions of any given mass may be brought

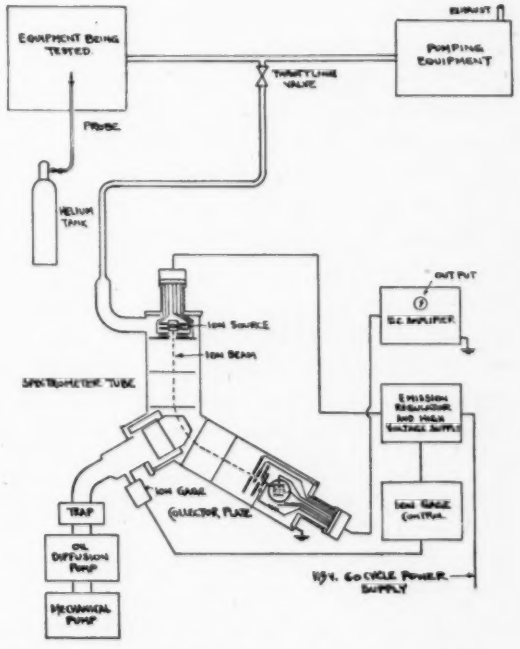


FIG. 1. Schematic line diagram showing use of mass spectrometer leak detector.

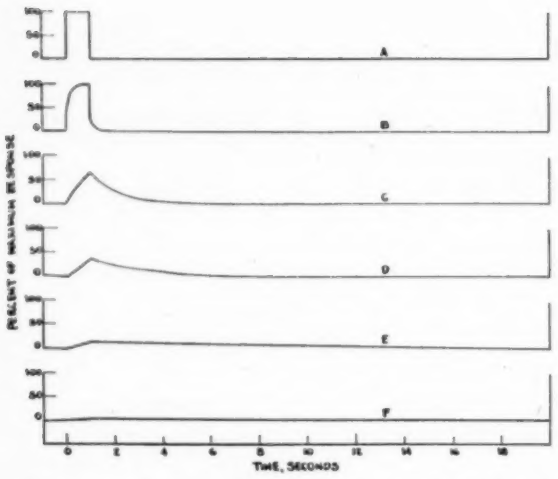


FIG. 2. Appearance and disappearance of probe signal for systems of varying response characteristics. (Probing is for a period of one second beginning at $T=0$.) Curve A is the probe pulse at the leak. Curves B through F have S/V values of 300, 60, 30, 6, and 3 minutes^{-1}, respectively, when S is expressed in efm and V as ft.^{3}.

into the collector by properly adjusting the electrical fields. For leak detecting purposes, the instrument is adjusted for He^+ ions, since helium is a most satisfactory probe gas. Theoretically, the instrument, when thus adjusted, will show a zero reading, unless helium is present in the gas stream under analysis. Actually, however, due to necessary physical imperfections in the apparatus, there is a continuous "background" reading. The presence of helium then is indicated by an increase over the normal background reading.

A separate report[1] gives detailed information on the construction and operation of the instrument developed at the University of Minnesota. Accordingly, we are concerned here, principally, with the performance of the instrument as a leak detector.

The first mass spectrometers used for leak testing purposes were all glass instruments using mercury diffusion pumps. The fragility of these machines for plant use was recognized, but the magnitude of the trouble caused, thereby, was discovered only after attempting to institute the production testing program. The entire spectrometer was redesigned by the Kellex Special Instrument Group. In addition to greater durability due to the practically all-metal construction (only the filament stem and the ion gauges remained glass), the new tube possessed a higher inherent sensitivity from design, and a greater sampling capacity through the substitution of an oil diffusion pump for the mercury model.

3.2. Dynamic Requirements for Rapid, High Sensitivity Vacuum Testing

Use of a high sensitivity, dynamic, leak detector such as the mass spectrometer is of little avail if the system under test does not possess the required characteristics. It is shown mathematically, in Appendix 7.4, that the ratio of S/V (pumping speed of the system with respect to the volume pumped) is of greatest importance, both with respect to dynamic sensitivity and speed of leak hunting. In Fig. 2, the response (at the leak detector) after exposing the leak to probe gas for one second is plotted for different pumping speeds.

The sharpness of the response to probing will, in a large measure, determine the ability of *any leak detector* to function efficiently. Figure 2 illustrates the impossibility of attaining a sharp response without a sufficiently high S/V ratio: i.e., about 6 or more reciprocal minutes.

When a large leak is probed, the system becomes temporarily "flooded" with a probe gas, and it is impossible to continue leak hunting until this gas is removed. In Fig. 2 probing is discontinued at the end of 1 second. Those parts of the curves for time greater than 1 second represent the "clean-up" period. The length of time required for "clean-up" is a function of the ratio S/V. Thus, for example, if a large leak is probed until a probe response of 25 percent of maximum is obtained, and pumping is then continued until all but 0.5 percent of the probe gas is removed, the "clean-up" times indicated in Table II are obtained. These calculations have been confirmed experimentally.

Loss of time by slow clean-up can prolong the leak hunting period considerably, perhaps much more than is indicated in Table II. The reason for this is the following: When one is hunting for leaks with too small a pump (low S/V), the dynamic sensitivity is low, and as a result, leaks

TABLE II. Evaluation of clean-up time as function of S/V.

S/V, minutes^{-1}	300	60	30	6	3
Clean-up time, seconds	1.7	5.8	11.1	139	546

must be carefully and repeatedly probed to establish their exact location. Consequently, by the time this is done, a considerable quantity of probe material will be introduced into the system. Since this must be removed before probing can continue, the leak detecting crew finds it necessary to remove a large quantity of probe gas at a slow rate, which is doubly time consuming.

A simple experiment was made which confirmed the above predicted results. The calculations have since been confirmed many times in practice, both under mass production and field conditions.

It must be recognized, that the S/V ratio refers to the speed of the entire vacuum system, including the process equipment under test. Thus the largest high speed vacuum pump will be of little use if the rest of the system throttles the flow to the pump. Rather, the use of a pump oversize for the system speed will further penalize the sensitivity, by making it impossible to obtain an adequate sample in the gas evacuation line through the leak detector. It is necessary that the lines of the system under test be sized to permit adequate speed of pumping. This can be done by sizing them for molecular flow according to the formula: Speed $= 14D^3/L$, where the speed is expressed in cubic feet per minute, D is the internal pipe diameter in inches, and L the length in feet. One of the more important functions of the Vacuum Engineering Division was to check all design to assure that the sufficient speed was always attained at the test connections.

The need for high speed pumps is apparent from the preceding discussion. These pumps had other special requirements:

1. The pumps and all associated vacuum equipment, i.e., valves, gauges, and piping, had to be tight, and engineered to maintain this tightness.
2. Provision to reduce the back diffusion of diffusion pump oil, without seriously impeding the speed characteristics of the pump, was necessary to prevent the contamination of the process equipment under test.
3. The complete pump unit, including mechanical and diffusion pumps, and all necessary controls, had to be self-contained and portable, and should require only connections to the system under test, water, power, and drain lines, to permit operation.

A typical vacuum pump unit is shown in Fig. 3. For initial evacuation, provision is made to by-pass the diffusion pumps through a roughing line. Pump units of suitable design were developed by Westinghouse Electric Corporation, National Research Corporation, and Distillation Products, Incorporated, in conjunction with the Kellex Vacuum Engineering Division.

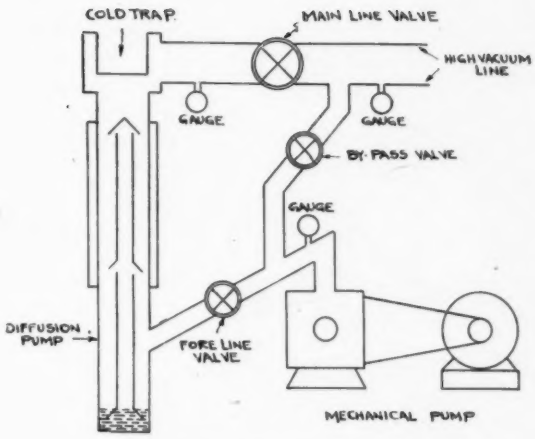


FIG. 3. Typical pumping unit.

4. MEASUREMENT OF TIGHTNESS

The general plan of attack, to secure over-all plant tightness, was to apportion the total allowable leakage among the different plant components, and to check individually for tightness each such component, one or more times prior to its installation.

This measurement of tightness, involving about 1,000,000 individual tests in all, threatened to be a serious bottleneck in the fabrication schedule. The only known method for measuring tightness consisted of evacuating the equipment sufficiently long to outgas it thoroughly and then to measure the pressure rise. This method usually required an average of 12 hours of continuous evacuation for each piece of equipment tested. Since the estimated number of vacuum tests on the individual components of the plant totaled over 600,000, it is evident that approximately 1000 leak rate stations each working more than 8000 hours would be required to prove the tightness of equipment before installation. This estimate does not include the time spent leak hunting, setting up the equipment for test, and retesting of rejects.

Evaluation of the mass spectrometer as the basic leak detection instrument indicated that its

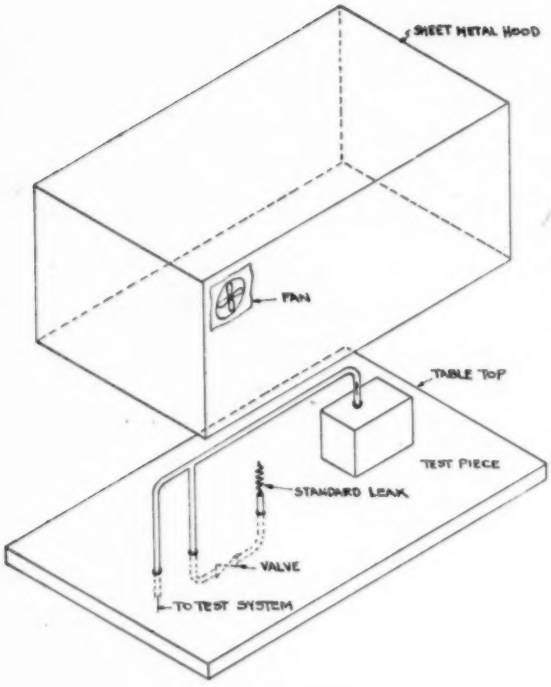


FIG. 4. Typical hood test set-up.

quantitative aspects could be adapted to measurement of the inleakage into the vacuum system. Accordingly, the so-called "Hood Method" for tightness testing was devised. Using this method in conjunction with the rapidly responding dynamic system for leak detection, the testing load for production testing was reduced to approximately 100 stations each working 2450 hours.

The Hood Method consisted of surrounding the equipment under test with a gas-tight hood in which a fixed concentration of probe gas was maintained, and comparing the amount of the probe gas entering the system through the leaks in the equipment to the amount of probe gas entering through a standard calibrated leak. This comparison was made by direct reading of the output of the mass spectrometer. The results thus obtained have been quantitatively checked against results obtained by measuring the inleakage of thoroughly outgassed vessels by the usual pressure build-up method.

A typical hood set-up is shown in Fig. 4. For small objects several manifolds are attached to one pump-leak detector combination, allowing the objects to be set up or taken down on one manifold, while the units on the other are under test. For larger equipment, individual pump-leak detector combinations are required. The test is made in the following manner:

With the valve leading to the calibrated leak in an open position helium is forced into the hood until its concentration there is sufficient to give a predetermined reading on the leak detector. This reading (reading 1) is proportional to the total inleakage consisting of: (a) unknown vessel leakage, (b) known leakage from the calibrated leak.

The valve leading to the calibrated leak is then closed, and a second reading taken (reading 2). This reading is, of course, proportional to the unknown leakage alone. From these two readings the following equation may be derived:

$$\text{Unknown leakage} = \frac{\text{Rdg (2)}}{\text{Rdg (1)} - \text{Rdg (2)}} \times \text{calibrated leak.}$$

This equation gives the unknown leakage in terms of the two readings referred to above and the value of the calibrated leak. Knowledge of the helium concentration in the hood is not necessary. The only requirement is that this concentration remains substantially constant during the time the readings are taken. Normally, this requires from five to ten minutes. Well constructed hoods show practically no decrease in helium concentration for considerably longer periods.

A modification of the hood technique, that has proved extremely useful for testing "hidden" leaks, involves the use of undiluted helium on both the standard leak and the unit under test. This method is best achieved by evacuation of the re-entrant interior volume suspected of

TABLE III. Contribution of various process components to inleakage (based on manufacturer's specifications).

Component	Volume contribution per 100 ft.³	Inleakage specification pressure build-up, microns/hr.	Inleakage contribution, mcfh
Diffusers	68.6	0.25	17.1
Centrifugal pumps	2.2	0.50	1.1
Process drums	3.8	0.1–1.0	0.4
Process piping and valves	25.4	1.00	25.4
Instruments	<0.1	<1.00	<0.1
Total per 100 cubic feet		0.44	44.0

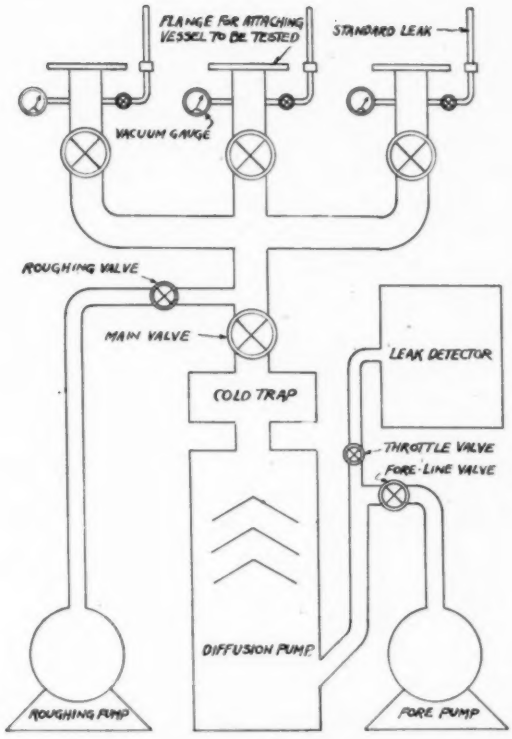


FIG. 5. Manifold for medium volume—medium production vacuum testing.

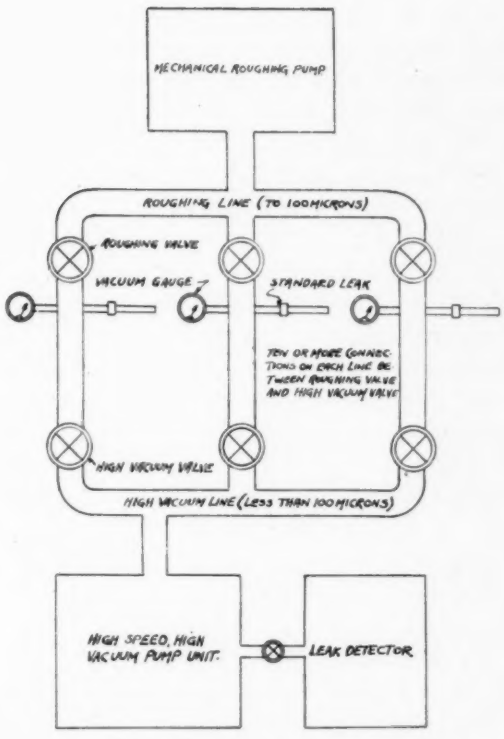

MECHANICAL ROUGHING PUMP
ROUGHING LINE (TO 100 MICRONS)
ROUGHING VALVE
VACUUM GAUGE
STANDARD LEAK
TEN OR MORE CONNECTIONS ON EACH LINE BETWEEN ROUGHING VALVE AND HIGH VACUUM VALVE
HIGH VACUUM VALVE
HIGH VACUUM LINE (LESS THAN 100 MICRONS)
HIGH SPEED, HIGH VACUUM PUMP UNIT
LEAK DETECTOR

FIG. 6. Manifold for small volume—high production vacuum testing.

leaking into the vacuum space proper, and breaking the vacuum with pure helium. Then the reading of the mass spectrometer output is compared to that obtained when a standard leak is surrounded by pure helium.

5. PRODUCTION TESTING

Experimental evidence indicated that the process specifications could probably be reduced from eight microns per hour to two microns an hour through the use of the methods just described. It was evident that most careful preassembly testing would be necessary to hold the number of leaks in the assembled plant to an absolute minimum. That is, it was most essential to be able to certify that various complex pieces of process equipment would not have to be disassembled for repair in the field, and that the vacuum testing crews in the field could have confidence in the tightness of the component pieces of equipment. The decision to pre-test every piece of process equipment and piping subassembly required the preparation of many varied procedures and specifications, and the education of manpower to use these new techniques in the different places of manufacture. Specifications for each unit were made on the basis of mechanical complexity and the number of such units occurring in each assembled cell of the plant. Table III lists the major components of each cell, the contribution of their volume per one hundred cubic feet of cell volume, and the inleakage contribution and the specified inleakage rate. It should be noted that, if no damage occurred during transport and erection, and if all field welds were tight, the assembled plant would have an inleakage rate considerably below 2 microns pressure rise per hour. Thus a margin was available for depreciation and safety.

Each manufacturer was asked to send several engineers to Kellex for training in the techniques of leak hunting, using the high sensitivity dynamic method and the hood test for the tightness testing. At the conclusion of the training program these engineers were asked to prepare detailed procedures applicable to their particular process unit. These procedures were reviewed by the Kellex Vacuum Engineering Division before they

were put into practice. In all cases high S/V ratios were maintained for production testing in keeping with the heavy production schedules. Many jigs and fixtures were designed and built for this purpose; since the undependability of rubber hose and similar makeshift connections was recognized, large sized equipment had flanged connections for testing, while for small equipment the Dresser type coupling proved successful. Best vacuum practice was followed in the construction of testing rigs; packless valves were used and all possible joints were welded, brazed, or hard soldered. In several instances, rigs were equipped with packed valves and sliding jigs; however, these units always proved to have lower sensitivity than those of all metal construction. Figures 5 and 6 show schematically some of the manifolds used for production testing.

6. TESTING THE ASSEMBLED PLANT

Delivering sound, vacuum-tight equipment to the plant site was one job. Perhaps a bigger one was to see that this equipment was assembled without injury, and that the assemblies were made tight. This program required the following groups: (1) Vacuum Engineering; (2) Construction Coordination; (3) Vacuum Testing; (4) Acceptance Control and Records.

Vacuum Engineering for the assembled plant was concerned with four major problems:

(1) Definition of the subsections of the plant to be tested as units.
(2) Location of test connections in process system, to assure sufficiently high S/V ratios for efficient testing.
(3) The magnified effect of sorption of the probe gas, when a large number of diffusers were tested simultaneously.
(4) Excessive extraneous leakage through the shaft seals of the centrifugal pumps.

Provision of sufficient test connections on the test cells allowed a careful study of all these problems to be made.

The process design of the plant largely determined the basic sections for testing, since the addition of valves to facilitate testing might interfere with orderly process operation. Thus a cell, building and cell by-pass lines, and the cold traps, were tested as subsections. The limits and state of completion of each subsection were accurately defined.

Each proposed layout of test connections was analyzed using an electrical analogy. The resistance of the piping to gas flow may be considered equivalent to electrical resistance. Consequently, an analysis of resistance network was used to demonstrate the optimum location for the test connections. These analyses were made mathematically or were measured on resistance networks. The results were quantitatively confirmed by later field measurements.

The Construction Coordination Group reviewed the progress of construction and established standards for completion of defined subsections. Briefly, these subsections were defined so that (1) no interference between construction and testing could exist (i.e., all subsections were ended past a limiting valve, so that opening of the valve to permit welding was unnecessary); (2) operations could follow vacuum testing so as to bring the process on stream in a practical orderly manner. Before transfer of buildings or sections to the Vacuum Testing Groups the Construction Coordination Group reviewed the unfinished construction items which might interfere with vacuum testing, and checked the results of preliminary pressure testing. Once a section of the plant was turned over to the Vacuum Testing Group, further construction made necessary by redesign or other reason was subject to the approval of the Construction Coordination Group.

Actual vacuum testing was performed by three-man crews assigned to each leak detector-pump unit combination. These crews were headed by a person of at least high school education and were all given a two-week special training course in the mechanics of the leak testing operation. Because of the complicated nature of the leak detectors and pump units, the test crews were not permitted to make any adjustments or repairs to the equipment. Highly trained personnel was used for the maintenance of all leak detection equipment.

Vacuum Testing Acceptance Control and Record Groups kept a record of all leaks found, the status of testing, and repairs. The results of studies by this group suggested several changes in design, erection procedures, and testing sequence, which considerably reduced both the probability of leaks and the time required for testing. The

major function of this group was to review the test data for conformance to specification.

The efforts of the Vacuum Testing Program involving well over 1100 men were climaxed when the plant was brought below a leak rate of 2 microns per hour. This was accomplished in the face of a constantly accelerated construction schedule and achieved within a very narrow margin of time allotted for vacuum testing after the completion of construction.

7. APPENDICES

7.1. Fundamentals of Vacuum Engineering*

Industrial uses of high vacuum have increased very rapidly in the past few years. Among such applications may be listed the manufacture of radio and x-ray tubes, lamp bulbs, rectifiers, etc., and the reduction of various metal bearing ores.

To satisfy the growing needs of industry in the field, much time and energy have been spent in the design and development of rugged, high capacity pumping equipment, by means of which the pressure in vessels of several hundred cubic feet of volume may be reduced to exceedingly low values with great rapidity.

(*a*). *Tightness and Cleanliness*

Tightness and cleanliness are absolutely essential in all vessels used in high vacuum systems. This becomes obvious when one considers that one cubic inch of standard air will occupy a volume of approximately 760,000 cubic inches at one micron pressure, and that one cubic inch of water when vaporized will occupy 945,000,000 cubic inches, at one micron pressure. Unless systems are clean and free from leaks, it is, therefore, extremely difficult to reduce the pressure in them to the high vacuum range.

(*b*). *Vacuum Pumping System*

A vacuum pumping system consists of three essential units known as the fore pump, the diffusion pump, and the vapor trap. (See Fig. 3.)

Diffusion Pump. The diffusion pump provides a means of obtaining large volumetric pumping capacities in the high vacuum range of pressures. The pumping action originates in the streaming of either oil or mercury vapor from a suitably designed jet. The gases to be pumped diffuse into this vapor stream and are carried along with it away from the system being evacuated. Pumps of this type usually attain their full volumetric capacity at about one micron pressure, and maintain this same value to very low pressures. The capacity of a well-designed pump will be approximately 50 cubic feet per minute per square inch of opening at the jet, and is, therefore, nearly $\frac{1}{3}$ the molecular flow through a hole equal in diameter to the pump housing when absolute vacuum is maintained on one side.

The diffusion pump will not exhaust directly against atmospheric pressure. There is a critical fore-pressure or threshold pressure above which it will not pump at all. This required fore-pressure (about fifty microns) is maintained by a mechanical pump, an ejector, or a combination of the two.

Fore Pump. The fore pump is a mechanical pump or ejector. It must have sufficient capacity to handle the gases compressed by the diffusion pump and maintain the fore pressure against which the diffusion pump operates at a value lower than the critical threshold pressure. Mechanical pumps will not in themselves produce pressures lower than about one micron. For lower pressures, the diffusion pump must be added.

Vapor Trap. The vapor trap is necessary to prevent the migration of oil vapor from the diffusion pump to the system being evacuated. In most cases, a suitably designed water-cooled baffle will suffice. However, a refrigerated cold trap is required to obtain pressures lower than 0.01 micron, or when it is desirable to prevent small traces of oil from entering the system which is being exhausted. Pressures lower than 0.01 micron may be obtained without the use of cold traps, but only with great difficulty and an excessively long "conditioning" time.

(*c*). *Flow of Gases*

The flow of gases in the high vacuum range obeys laws quite different from those familiar to the hydraulic engineer. This fact cannot be stressed too highly since it is overlooked all too frequently by the uninitiated.

Extremely large and short pipes must be used to handle the large capacities of a diffusion pump,

* By J. R. Downing, Kellex Corporation. (Present address: Cook Electric Company, Chicago, Illinois.)

TABLE IV. Capacity loss through pipe connections to a diffusion pump (capacity 2600 cfm below 0.1 microns).

Length of pipe in feet	Percent capacity reduction		
	4" I.D.	8" I.D.	12" I.D.
2	88	52	27
6	95	73	45
20	98	89	71

although the resultant mass transfer is very small. For piping systems longer than a few feet, the pipe size used must be of larger diameter than the pump opening. It is customary to calculate pumping system capacities in terms of the capacity of the pump and the flow capacity of the pipe. The expression for over-all pumping capacity is:

$$\frac{1}{S_{\text{system}}}=\frac{1}{S_{\text{pump}}}+\frac{1}{S_{\text{pipe}}}$$

where the various pumping capacities are expressed in cubic feet per minute. If the length of the pipe is more than ten times the diameter, the capacity of the pipe for air is given by:

$$S_{\text{pipe}}=\frac{0.51D^4P}{L}+\frac{14D^3}{L}\frac{1+0.62DP}{1+0.76DP},$$

where D is the pipe diameter in inches, L is the length in feet, and P is the average line pressure in microns. A second semi-empirical expression for the capacity of short lengths of pipe at pressures less than 0.1 microns is given by:

$$S_{\text{pipe}}=\frac{14D^3}{L+(D/9)},$$

where L, D, and S have the same units as in the above formula. It is evident from these expressions that the pipe size may be the limiting factor in the capacity of a pumping system regardless of the capacity of the pump. Table IV gives the reduction in capacity caused by various lengths of pipe connecting a 2600 cfm diffusion pump to the vessel being evacuated.

(*d*). *Precautions and Techniques*

Precautions and techniques necessary to the attainment and maintenance of high vacuum are so many and varied that the subject will be dealt with here only in the most general way. This is summed up by saying that vacuum systems must be capable of attaining high vacuums without the use of waxes, greases, lacquers, soft solders, and the like.

Sound engineering, cleanliness, ruggedness, and good materials are all of prime importance.

The equivalent in weight and strength of standard 150-pound flanges and fittings is recommended. These may seem excessively heavy but, in view of their reliability and trouble-free operations, they are well worth the additional cost.

(*e*). *Evacuation Time*

The time required for evacuation may be calculated fairly well for the range of pressures from atmospheric down to a few hundred microns. The expression is:

$$T=\frac{V}{S}\ln\frac{P_1-P_0}{P_2-P_0},$$

where T is time in minutes, V is volume of vessel in cubic feet, S is the pump capacity in cubic feet per minute, P_1 is the initial pressure, P_0 is the pressure attained by the pump above at blank off, and P_2 the final pressure. Since P_0 is small compared to P_1 and P_2, the expression may be simplified to:

$$T=\frac{V}{S}\ln(P_1/P_2).$$

With P_1 atmospheric pressure and P_2 a few hundred microns, the term involving the logarithm becomes about eight so that the time required for "roughing" a vessel becomes:

$$T=8(V/S).$$

The latter expression holds quite well, provided the system is free from inleakage and the volumetric capacity (S) of the pump stays reasonably constant. For most mechanical vacuum pumps the latter condition is true down to pressures of a few hundred microns.

In the range of pressures lower than 100 microns, the above expression is quite useless for three reasons:

1. The capacity of the diffusion pump is by no means independent of pressure.

2. Degassing and vaporization of substances on the inner walls of a vessel contribute so greatly to the volume of gases which must be handled by the pump, that the residual volume of air becomes negligible by comparison.

3. Impedance of connecting pipes becomes an important factor.

If the amount of degassing per unit surface area is known as a function of time and pressure, then the pump-down time may be estimated by a graphical integration method. However, such data are usually not known very accurately so that choice of pump size is dictated largely by experience. The modern trend is to use diffusion pumps of very large capacity so that a high vacuum may be produced quickly and in spite of degassing, small leaks, etc.

(*f*). *Special Gauges*

Special gauges have been developed for the measurement of pressures lower than one millimeter of mercury. The forces involved are so small that a direct measure of the pressure by observing the deflection of a diaphragm or the equivalent does not prove practical. Some other property of the gas which varies with pressure and which is capable of easy application must be chosen.

The McLeod gauge has long been the standard for the accurate measurement of pressures less than one millimeter of mercury. A sample of the gas in the system is entrapped and compressed by a known ratio. A measure of the resulting pressure then permits calculation of the initial pressure. Compression ratios as high as a million may be used enabling the measurement of pressures as low as 0.001 microns. Because of the high compression ratio, the gauge may only be used for the measurement of pressures due to permanent gases.

The thermocouple and Pirani gauges depend upon the variation of heat conductivity with pressure of the gas. These gauges are, therefore, not absolute and must be calibrated by comparison with the McLeod gauge described above. They are sensitive in the range of pressures between one millimeter of mercury and one micron.

The ionization gauge is capable of measuring pressures from one micron to the lowest attainable which, at present, is about one millionth of a micron. The gauge is essentially a three-element vacuum tube operating with its grid at a positive potential and plate negative with respect to the filament. Electrons from the filament gain energy in their passage to the grid and so will cause ionization of those gas molecules with which they may collide. The resulting positive ions are collected at the plate and constitute the plate current. Since the probability for ionization is proportional to the molecular concentration of the gas, the plate current represents a measure of the pressure.

Ionization gauges, like the thermocouple or Pirani gauges are not absolute, but must be calibrated by comparison with a McLeod gauge.

A number of other gauges have been developed for measuring low pressure, but few if any of them have ever proved practical in industrial applications. There is considerable room for improvement in apparatus for measuring low pressures, since none of the gauges now in use are wholly satisfactory.

7.2. Pressure Testing Techniques

Pressure testing for tightness has been applied to vessels in both vacuum and pressure service for many years. Therefore, this section of the report will merely indicate the various methods briefly and discuss their applicability to testing process equipment for high vacuum service. Table V indicates the various methods in general use and indicates the smallest leak that may be detected practically. Extremely careful application of the method will perhaps decrease the quoted figure two- to fivefold.

TABLE V. Pressure testing techniques.

Test medium	Probe	Response	Leak size at 45 psig, mcfh	Special requirements
Air-nitrogen	Flame	Wavering	5000	Draft free room.
	Sound	Hissing	5000	Low noise level.
	Soap film	Bubbles	5	Each area to be observed under good light for at least five minutes and soap film maintained.
	Immersion	Bubbles	20	Five to fifteen minutes observation—occluded air on surface must be avoided.
Organic halide	Halide torch	Flame color change	5	
Acidic gas—(CO_2 or SO_2)	Ammonia	Fumes	5	
Ammonia	CO_2; HCl	Fumes	5	
Liquid under high pressure		Wet exterior surface	500	

These techniques are used only to locate leaks and do not permit any measure of the amount of leakage to be made. Application of pressure drop measurements are not, in general, suitable to checking the tightness of systems for vacuum service, the permissible drop being beyond the limits of accuracy of the best commercial gauges generally available. This drop for diffusion plant standards would be less than 0.001 lb./sq. in./hr. and corresponds to a temperature change of about 0.005°F. The uncertainty of the temperature correction further invalidates this method for measurement of the vacuum tightness of process vacuum equipment.

After all leaks are thought to have been found, the equipment under test must be evacuated and the rate of pressure rise observed. Consideration of the gas laws reveals that the temperature correction in the absolute pressure range of 1–100 micron is negligible. Vacuum gauges can readily measure the pressure changes required for tightness. However, a serious drawback is the effect of outgassing on the observed leak rate. The subject is more completely discussed in Appendix 7.3.

Other practical drawbacks to the use of pressure testing techniques are (1) the requirement that the system be depressured before repairs are attempted; (2) inability to check hidden leaks (leaks that are not visible on the exterior to the system) such as valve seats and heat exchanger tubes.

7.3. Vacuum Testing Techniques Prior to 1943

Many devices used in vacuum practice will not stand internal over-pressures high enough to make pressure testing practical; therefore the use of vacuum testing techniques became necessary in the laboratories and specialty shops dealing with high vacuum equipment. These methods were considered usable only by specialists, and were not considered applicable to commercial process equipment before the inception of the testing program described in this report. The earliest form of vacuum testing was on all glass systems. After evacuation of the system was begun, the discharge of a Tesla coil was played over the glass surface. A leak would show up as a bright spot or "pip." Experienced operators could detect leaks on metal parts of the system by observing the color change of the discharge when the leaks were sprayed with water or acetone. However, considerable uncertainty was attached to this technique, and one often resorted to "painting" the metal system with a vacuum wax, or lacquer, on a wiped layer of solder or lead.

Perhaps the first large scale vacuum testing of metal systems was done by research groups working with cyclotrons and similar equipment used in nuclear research. In general, those methods depend upon the effect on a hot wire gauge of the change of composition of the gas in the system when the probe fluid displaced the air flowing through the leak. Precautions were taken to insure electrical stability and sensitivity of the measuring circuits and mechanical stability of the gauges used. Probe fluids were carefully chosen to insure maximum effect on the gauge. Thus the ideal probe fluid had low viscosity and either high or low thermal conductance compared to air.

When tightness requirements became more severe, the hot wire gauge was replaced with an ion gauge, since the tightness was sufficient to permit the evacuating pumps to reduce the pressure below the operating maximum of this gauge (<0.5 micron). Hydrogen gas or a gaseous hydrocarbon was most generally used for the probe fluid, although good results were reported for other fluids. Ease of ionization of the probe molecules is a determining factor in the choice of the probe fluid, since the ionization produced by electron bombardment is proportional both to the concentration of the gas in the gauge and to the character of the gas.

To obtain the ultimate sensitivity of these methods the dynamic method was discarded and one of the following methods adopted: (1) the "backing space" technique, or (2) the apparent change in rate of pressure build-up when probe gas displaces air flowing through the leak. Both methods achieve the improved sensitivity at the expense of an increase in testing time.

Backing space is the name given to the process of multiplying the effect of the probe gas by permitting the diffusion pump to compress the gas into a dead volume. This is achieved by shutting the valve in the foreline leading to the mechanical pump. The method further requires judgment in the analysis of the signal indicated on the gauge controls. The caliber of the personnel required to assure foolproof operation is quite

TABLE VI. Resumé of vacuum testing methods in use prior to 1943.

Leak detector	Operating range, microns	Size of leak discernible, mcfh	Commonly used probes	Remarks
Tesla coil	50–1000		Tesla coil	Can be used only on glass systems.
Discharge tube	100–1000		Acetone; Methanol; CO_2; H_2	Residual gases confusing.
Hot wire gauge	Less than 100	10–1000	Acetone; Methanol; H_2	Affected by pressure changes and residual vapors.
Ionization gauge	Less than 0.5	1–100	Gaseous hydrocarbons; H_2; O_2	Affected by pressure changes and residual vapors.
Hot wire with backing space	Less than 100	50–100	Gaseous hydrocarbons; CO_2; H_2	Time-consuming since method is discontinuous.
Ionization gauge with backing space	Less than 0.5	1–100	Gaseous hydrocarbons; CO_2; H_2	Time-consuming since method is discontinuous.
Hot wire with pressure build-up	Less than 50	1–100	Gaseous hydrocarbons; H_2; CO_2; masking with vacuum putty	Extensive outgassing required.
Ionization gauge with pressure build-up	Less than 0.5	1–100	Gaseous hydrocarbons; H_2; CO_2; masking with vacuum putty	Extensive outgassing required.

high—this method having been reported by groups using only post-collegiate personnel.

Analysis of the apparent change in pressure build-up caused by the introduction of probe gas is very time consuming, requiring that a careful schedule for probing be adhered to, and a subsequent analysis of the time-pressure graph in conjunction with this schedule be made. A further drawback of the technique is the entirely nonproductive testing time spent reducing the outgassing of the system under test. This period varies between 8 to 48 hours per unit depending on the volume and area of outgassing surface involved.

Having found all possible leaks, a measure of tightness is taken in the following manner. The equipment under test is valved off from the evacuating pump and the pressure rise of the system observed. Evacuation of the equipment during the leak hunting period in general has reduced the outgassing sufficiently so that the pressure rise is an approximately true measure of the inleakage. Outgassing may be detected on the pressure-time graph of the leak rate by a pronounced change of slope. The true leak is generally taken after this change of slope has been noted. The rate thus measured does not differ appreciably from the rate of the same system after it has been completely outgassed.

Outgassing is a complex and unpredictable effect, as the gas contributing to the effect may be adsorbed, absorbed in solid or liquid solution, evaporated from a liquid, or the product of chemical decomposition.

A resumé of vacuum testing methods prior to 1943 is presented in Table VI.

7.4. Relation between the Volume Pumped, Pumping Speed, and the Response to the Probe

Assumed that L standard cubic feet per minute are leaking into a volume of V cubic feet which is being pumped at a rate of S cubic feet per minute at the total pressure existing in V.

Then, if at time $t=0$ probing of the leak is initiated, the volume V will:

(1) Gain: Ldt (standard cubic feet of probe gas per differential of time dt).

(2) Lose: $pSdt$ (standard cubic feet of probe gas per differential of time dt) where p is the partial pressure of probe gas, expressed in atmospheres, in V.

The net gain of probe gas in volume V is:

$$d(pV)=(L-pS)dt,$$

or

$$V\frac{dp}{dt}=(L-pS).$$

Integration of this differential equation gives the partial pressure of the probe gas in the volume and accordingly the partial pressure, p_t, of the probe gas in the stream passing the leak detector at any time t after probing of the leak has begun:

$$p_t=\frac{L}{S}[1-\exp(-St/V)].$$

If, at time T, the probing is discontinued, a similar analysis will show that the partial pressure of the probe gas remaining in the system will be given by the equation:

$$p_T=\frac{L}{S}[1-\exp(-ST/V)]\exp[-S(t-T)/V].$$

In Fig. 2 the response and clean-up to a probe applied to a leak into a volume of forty cubic feet for a period of one second is shown. The response is expressed in terms of percent of the equilibrium response which is given by the equation:

$$p_\infty=L/S.$$

The Magnetic Electron Microscope Objective: Contour Phenomena and the Attainment of High Resolving Power

JAMES HILLIER AND E. G. RAMBERG
RCA Laboratories, Princeton, New Jersey
(Received September 3, 1946)

A theoretical and experimental investigation of contour phenomena observed in electron microscope images near focus has been carried out. An explanation for the apparent reversal of the Fresnel fringes at exact focus is given and a method is described for the determination of the degree of asymmetry from the fringe patterns. A procedure for empirically correcting the asymmetries usually present in magnetic electron microscope objectives is outlined and some of the results obtained with a compensated lens are shown. A number of instrumental defects which may prevent the attainment of the ultimate resolving power of an instrument, together with methods for their elimination, are listed.

I. INTRODUCTION

IN a recent paper[1] it was pointed out that investigations made on the pole-piece system of the objective lens in a magnetic electron microscope operating under optimum instrumental conditions indicated that the practical limit of resolving power of this instrument had been imposed in the past by asymmetries in the lens field. The work also indicated that it may be possible to attain, in practice, the theoretical limit of the existing lenses if those asymmetries could be completely removed.

The present paper is intended as a more complete report on those studies and on others, made subsequently, which have led to a better understanding of the magnetic electron microscope objective, to quantitative methods of testing the quality of a given pole-piece system, and to a method of empirically compensating for the asymmetries in the imaging system of a magnetic electron microscope.

The work has indicated that the larger magnetic electron microscopes in use in this country are inherently capable of resolving powers of the order of 10A when provided with objective lenses of sufficiently high quality. Furthermore, the attainment of such a high level of performance appears to present a parallel to the attainment of the limiting resolving power in the light microscope in that it depends on the skill and experience of the individual microscopist as much as on the availability of particularly good lenses and other equipment. In fact, as will be seen from the following, the methods by which this performance has been achieved in this laboratory

[1] J. Hillier, "Further improvement in the resolving power of the electron microscope," J. App. Phys. **17**, 307–309 (1946).

are, for the present at least, not dependent on any commercial development and can be repeated in any electron microscope laboratory possessing normal facilities. Because of this, it is the further intention of the authors to present, in as detailed form as possible, procedures and precautions which must be observed in order to achieve the maximum attainable resolving power.

After the instrument used in this laboratory[2] had been "tuned up" to the point where no instrumental defects other than those inherent in the lenses were limiting the performance, at least at the 20A level, it was realized that asymmetries in the field of the objective were the most important cause of the remaining defects in the image and that the appearance of the image contours in the best-focused images was the most sensitive test of these asymmetries. Owing to the refractive effects in a transparent membrane, as discussed below, there is an apparent reversal of the Fresnel fringes as the power of the objective is varied through the focal value. In the best-focused images produced by a lens which is not symmetrical, that is, a lens for which the focal length is not independent of the azimuth of the meridional plane in which any particular paraxial ray is traveling, the reversal in the Fresnel fringes produces a characteristic type of asymmetry which is shown in Fig. 1. This is a high magnification image of an opening in a thin collodion membrane. Here, in one direction (vertical) the contours are those obtained when the power of the objective is less than the focal value, while the contours in the other direction (horizontal) are those obtained when the power of the objective is greater than the focal value. In Figs. 2 and 3, which are images of the same field as in Fig. 1, taken with the same lens, but with objective powers, respectively, much less and much greater than the focal value, the image asymmetries are no longer visible. In practice, the range of the objective current over which the image asymmetries are detectable provides a quantitative measure of the degree of asymmetry existing in the lens. As a further check it is

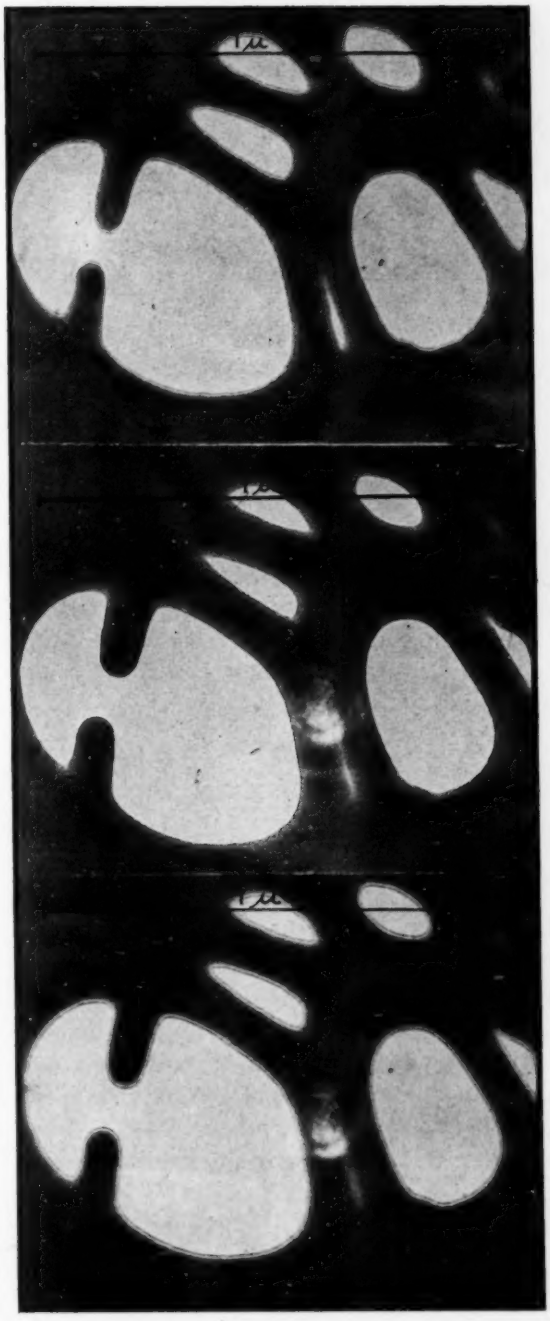


FIG. 1 (upper). Focused image of holes in collodion membrane obtained with asymmetrical objective. Vertical edges show fringes characteristic of under-focused image. See Fig. 2. Horizontal edges show fringes characteristic of overfocused image. See Fig. 3.

FIG. 2 (middle). Underfocused image (focal length too long) obtained with the same objective.

FIG. 3 (lower). Overfocused image (focal length too short) obtained with same objective.

[2] An RCA EMU electron microscope using the saturated emission type electron gun and fitted with a telescopic image viewing device (reference 1). With the exception of these differences and those introduced later as a result of the present work, the instrument is a standard commercial model.

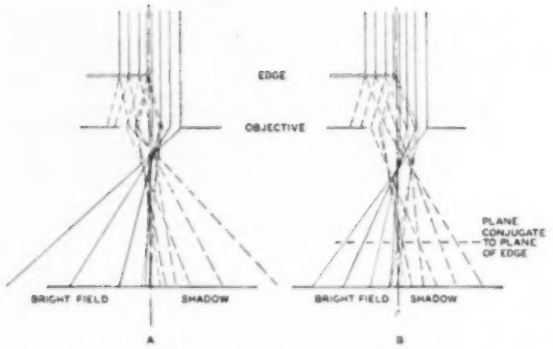


FIG. 4. The imaging of an edge by a parallel beam of electrons: A. Effect of spherical (or chromatic) aberration; B. Effect of defocusing.

possible to estimate the variation of focal length with the azimuth of the meridional plane from a photometric study of the asymmetries in the contours as described below. While important in the academic study of contour phenomena, this latter technique is cumbersome in practice and was not used in the actual compensation of the lenses.

II. THEORY OF CONTOUR PHENOMENA FOR SMALL ANGLE OF ILLUMINATION

At the edge of an object projecting into space or of a hole in a film the electron intensity distribution in the image ceases to bear a simple relation to the scattering power of the object. Fringe systems are observed whose intensity distribution depends both on the focusing of the objective lens and the condition of illumination. It will here be assumed that a straight edge intersecting the optic axis of the electron microscope at right angles is being observed and that the object is illuminated by a parallel beam of electrons.

Disregarding, to begin with, diffraction phenomena, it is seen that all the electrons passing through the unobstructed part of the object plane fall on the corresponding half-plane of the image, forming there a near-uniform intensity distribution, regardless of the condition of focus and the optical quality of the objective (Fig. 4). The electrons passing through the material portion of the object will be scattered to some extent, so that, except for the position of perfect focus of a perfect lens, a fraction of the electrons leaving the material portion of the object near

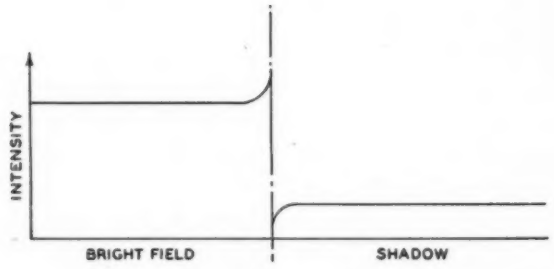


FIG. 5. Intensity distribution at the image of an edge illuminated by a parallel beam, neglecting diffraction effects.

the edge will fall on the bright half-plane of the image. Thus, for parallel illumination and with diffraction phenomena disregarded, spherical aberration, chromatic aberration, and defocusing all give rise to the same type of contour phenomenon—a bright fringe on the bright side of the edge and a dark fringe, antisymmetric to the bright fringe, on the dark side of the edge (Fig. 5).

The contour fringes normally observed with small angle of illumination are much more complex than this. This complexity must be ascribed primarily to Fresnel diffraction at the edge. The intensity distribution on a screen placed a distance l behind an opaque object illuminated by a parallel beam of light is shown in Fig. 6. It may be regarded as the result of interference between the unobstructed plane wave, cut off sharply at the geometrical shadow of the edge,

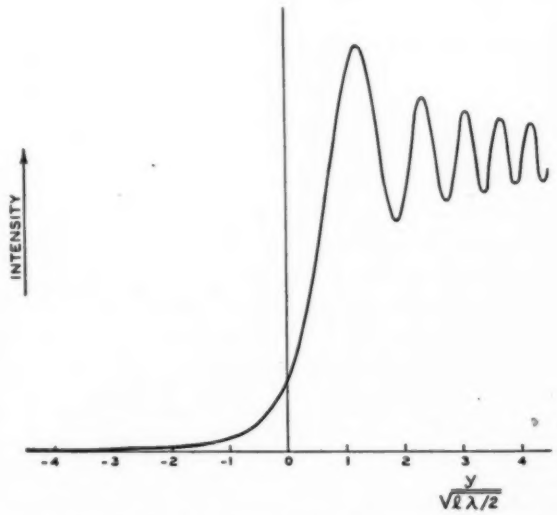


FIG. 6. Intensity distribution behind the edge of an opaque screen illuminated by a plane wave.

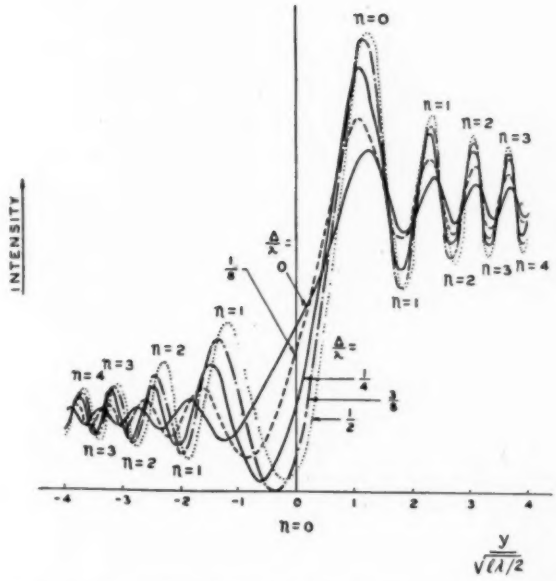


FIG. 7. Intensity distribution behind the edge of a screen transmitting a quarter of the incident radiation with a phase delay $2\pi\Delta/\lambda$ (real pattern).

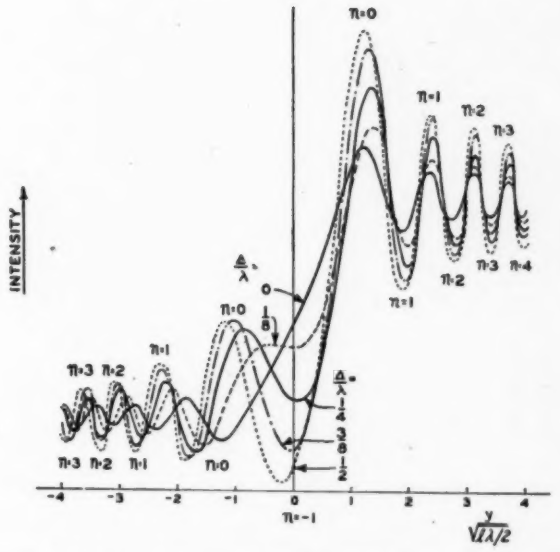


FIG. 8. Virtual intensity distribution ahead of the edge of a screen transmitting a quarter of the incident radiation with a phase delay $2\pi\Delta/\lambda$.

and a cylindrical wave diverging from the edge.[3] The intensity of the edge wave is a maximum in a forward direction and drops off laterally. Its phase is retarded by $\frac{1}{8}$ cycle with respect to the plane wave in the shadow of the edge, advanced by $\frac{3}{8}$ cycle in the bright region; at the shadow of the edge (where this representation ceases to be valid) a phase reversal takes place. In view of these relationships the maxima and minima of the diffraction pattern in the bright part of the field are located at a distance from the shadow of the edge given by

$$y_{\max}=(2l\lambda[n+\tfrac{3}{8}])^{\frac{1}{2}},\quad y_{\min}=(2l\lambda[n-\tfrac{1}{8}])^{\frac{1}{2}},\qquad n=0,\,1,\,2,\,3,\,\cdots. \tag{1}$$

Here λ represents the wave-length of the radiation employed. In the shadow the intensity falls off uniformly.

If a plane ahead of the edge is imaged on a screen by a perfect lens, an identical pattern is observed, though now the fringes result from the interference of the shadow portion of the edge wave with the direct wave. Since the wave advances in phase by a quarter cycle in passing through a line focus[4] the phase of the virtual wave continued backwards from the shadow portion of the edge wave is retarded by $\frac{3}{8}$ cycle with respect to the plane wave. Furthermore, the sign of the path differences for the two interfering waves is reversed in the virtual pattern, so that its maxima and minima have the same location relative to the shadow of the edge (continued backwards) as the maxima and minima of the real pattern. Mathematically, if the amplitude of the wave is given in complex notation, the amplitude in the virtual field ahead of the edge becomes simply the complex conjugate of the amplitude in the real field, an equal distance behind the edge.

If the screen transmits a given fraction p^2 of the incident radiation without destroying its coherence or modifying its phase, the expected pattern may be obtained from a superposition of the amplitudes for two plane waves of equal phase differing in initial amplitude by the factor p and incident on complementary opaque screens. Again the expected intensity distribution in a real plane behind the edge and the virtual distribution in a plane at the same distance ahead

[3] See, e.g., A. Sommerfeld, "Theorie der Beugung" (especially p. 843) in P. Frank and R. v. Mises, *Die Differentialgleichungen and Integralgleichungen der Mechanik and Physik* (Vieweg, Braunschweig, 1935), part 2, second edition.

[4] For derivation of this "Gouy effect" see P. Debye, "Light near a focal point or line," Ann. d. Physik **30**, 755–776 (1909).

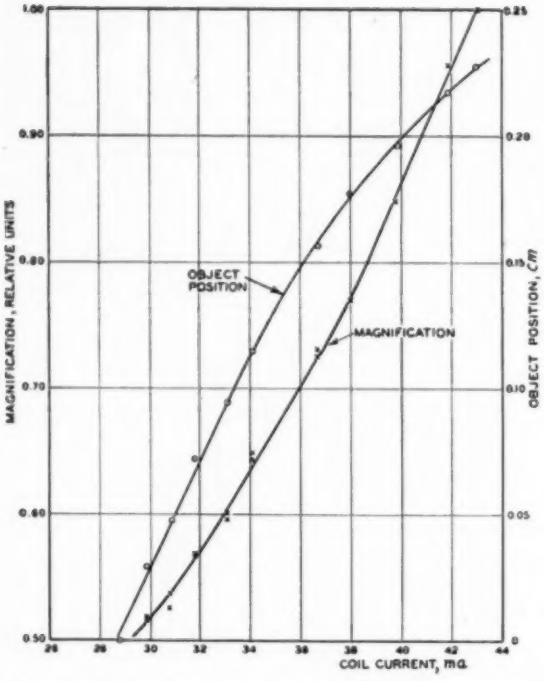


FIG. 9. Measured variation of object position and magnification as function of objective coil current.

of the edge are identical (Figs. 7 and 8, curves for $\Delta/\lambda=0$). It should be noted that in the real pattern the fringe system in the shadow region results from the interference of the wave transmitted through the object with the shadow portion of the edge wave, while in the virtual pattern this fringe system results from interference of the transmitted wave with the part of the edge wave scattered into the bright field.

If the screen with the transmission factor p^2 introduces a phase delay in the transmitted wave, the real and virtual diffraction patterns cease to be identical. Let, e.g., the difference in the index of refraction of the screen material and that of surrounding space be δn and the thickness of the screen be t; then the optical path difference between the wave passing through the screen and that passing through the clear space outside of the screen is $\Delta=t\delta n$. Thus, the phase delay introduced by the screen becomes $2\pi\Delta/\lambda=2\pi t\delta n/\lambda$. The intensity distribution in the diffraction pattern may now be obtained from a superposition of the amplitudes for two plane waves differing in initial amplitude not by p, but by the complex quantity $pe^{2\pi i\Delta/\lambda}$, incident on complementary opaque screens. This intensity distribution was calculated, both for the real and for the virtual pattern, for $p=0.5$ and $\Delta/\lambda=\frac{1}{8}$, $\frac{1}{4}$, $\frac{3}{8}$, and $\frac{1}{2}$. The results of the calculations are shown in Figs. 7 and 8. For $\frac{1}{2}<\Delta/\lambda<1$ the real and virtual patterns are identical with the virtual and real patterns, respectively, for $1-\Delta/\lambda$, so that, in effect, the curves in these two figures represent all possible intensity distributions for phase delays which are integer multiples of $\lambda/8$.

The results of the calculations may be summarized as follows:

1. A phase delay enhances the prominence of the diffraction fringes; it becomes greatest for $\Delta/\lambda=\frac{1}{4}$.

2. In the range $0<\Delta/\lambda<\frac{1}{2}$ the real pattern contracts to an increasing extent. If the fringes are counted off in the same manner as for the opaque screen (Eq. (1)), the fringes of high order number on the shadow side are displaced, for $\Delta/\lambda=\frac{1}{4}$, by half a fringe toward the edge; on the bright side they are not displaced.

3. For $\frac{1}{8}\leqslant\Delta/\lambda<\frac{1}{2}$ the virtual patterns possess an added minimum and maximum, which may properly be designated as the minimum of order -1 and the maximum of order 0 on the shadow side. The high order fringes on the shadow side are displaced a fraction of a fringe away from the edge.

III. OBSERVED CONTOURS

Contour fringes in out-of-focus electron microscope pictures have been observed and interpreted as Fresnel diffraction fringes by a number

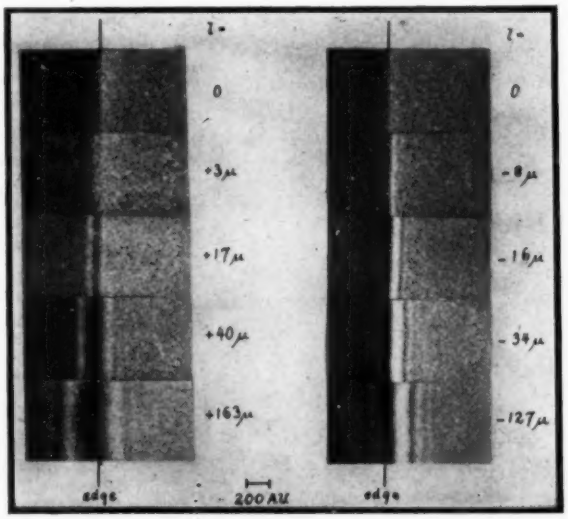


FIG. 10. Electron micrographs of the edge of a collodion film for various degrees of defocusing. The amount of defocusing is given to the right of edge section. Positive and negative values are for over and underfocused images, respectively. The true position of the edge is marked. Compare with Figs. 2 and 3.

of authors.[5–7] Boersch,[7] who made a particularly careful study of the phenomenon, noted that the fringes in the bright field were much more marked in the virtual pattern (objective lens too weak) than in the real pattern (objective lens too strong). He interpreted this as signifying that the shadow portion of the edge wave is much stronger than the edge wave scattered into the bright field. He had reached the same conclusion before from dark field observations.[8] These showed that object edges facing the direction of illumination appeared less bright than the edges remote from the source.

The observations to be presented here were made largely in the hope of shedding more light on the behavior of the fringes near focus. A collodion film with a number of holes of the order of 1 micron in diameter served as object. The illuminating cone was slightly elliptical and had maximum and minimum half-angles of 6 and $4 \cdot 10^{-5}$ radian, respectively. These were calculated from the observed image of the effective source, the condenser current, and the geometrical dimensions of the electron microscope which was employed. The variation in position of the object plane of sharp focus with the objective lens current was determined by displacing the object by measured amounts and reading the currents required to bring the image into sharp focus (Fig. 9). It should be mentioned that hysteresis effects in the lens pole pieces cause a variation in the relation between the object displacement and the lens current, so that possibly more accurate estimates of the former may be obtained from the observed diffraction pattern as indicated below. The object displacements as determined by the two methods differed randomly by 10 to 20 percent. The geometrical position of the edge relative to the pattern could be determined roughly, at large deviations from focus, by measuring the magnification of the plate and comparing the known diameter of the hole obtained at focus with the diameters of specific fringes about the hole edges. Results so obtained agreed within experimental error with the edge positions inferred from a comparison of the observed diffraction pattern with that predicted theoretically.

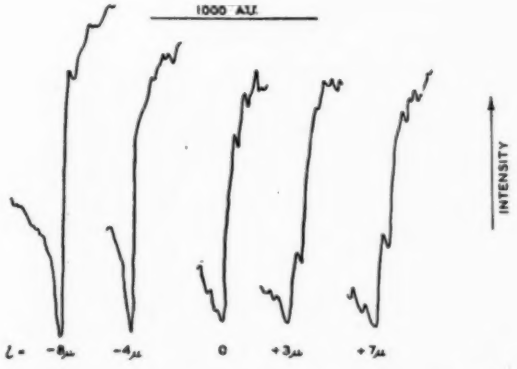


FIG. 11. Microphotometer traces of micrographs of edge taken near focus.

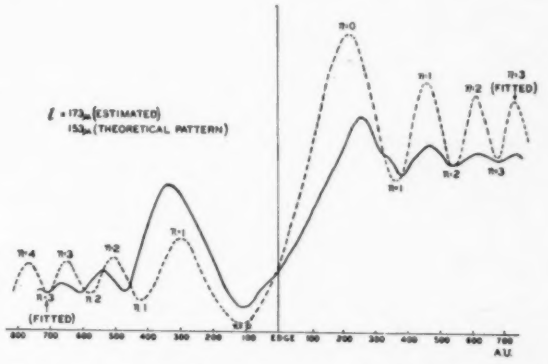


FIG. 12. Comparison of microphotometer trace and Fresnel pattern for edge of screen transmitting a quarter of incident radiation with a quarter cycle phase delay (real pattern).

Figure 10 shows a series of fringe patterns observed for positive (real pattern) and negative (virtual pattern) values of the deviation from focus, as indicated next to the individual micrographs. A vertical line marks the position of the edge. Figure 11 reproduces some microphotometer traces across the pattern for the region close to focus. An inspection of the contour patterns reveals the following:

1. Patterns corresponding to numerically equal deviations from focus in the direction of too great lens currents (real pattern) and too small lens currents (virtual pattern) differ markedly.

2. Proceeding from what, rather arbitrarily, will be

[5] E. Ruska, "The origin of fringes about particles in electron microscope images and their change with focus," Kolloid Zeits. **105**, 43–52 (1943).

[6] A. Prebus, "The electron microscope," p. 170, in J. Alexander, *Colloid Chemistry* (Reinhold Publishing Corporation, New York, 1944), Vol. 5.

[7] E. Boersch, "Fresnel diffraction in the electron microscope," Physik. Zeits. **44**, 202–211 (1943).

[8] H. Boersch, "Edge diffraction of electrons," Physik. Zeits. **44**, 32–38 (1943).

referred to as the condition of sharp focus ($l=0$ in Figs. 10 and 11) toward smaller lens currents, an increasingly sharp bright fringe appears on the bright side of the edge, followed by a dark fringe on the dark side. As the lens current is decreased still further, these fringes gradually move apart and added maxima and minima are resolved outside of either.

3. Proceeding from the point of sharp focus to higher lens currents, there appear at the edge a dark fringe and a bright fringe, the bright fringe lying on the shadow side of the dark fringe. These increase rapidly in prominence. Eventually fringe systems become visible outside of these primary fringes, as in the case of too low lens currents. It should be noted that the fringes in the shadow of the edge are much more prominent than in the case of too low lens currents. This is in accord with Boersch's observation that the amplitude of the part of the edge wave scattered into the shadow is greater than that of the part scattered into the bright field. For too high currents (real pattern) the shadow fringes are produced by interference between the transmitted wave and the shadow portion of the edge wave, for too weak lens current (virtual pattern), by interference between the transmitted wave and the portion of the edge wave scattered into the bright field.

IV. COMPARISON OF OBSERVATIONS AND THEORY

The presence of diffraction fringes in the shadow of the collodion film indicates that the latter has not introduced large *random* phase variations in the electron wave transmitted by it. On the other hand, a comparison of the microphotometer traces of the fringe systems for large deviations from focus with the curves in Figs. 7 and 8 indicates that a satisfactory agreement between the positions of the maxima and minima in the observed and theoretical patterns can only be obtained if the wave passing through the film is assumed to experience a phase delay. The curves for $\Delta/\lambda=\frac{1}{4}$ resemble the observed patterns most closely. Since $\Delta=t\delta n$ and $\delta n=\delta V/(2V)$, where t is the thickness of the film, δV the difference in the mean potential within the film and that outside of it, and V is the accelerating voltage of the electron beam, this leads, for $\delta V=10$ volts and $V=50{,}000$ volts, to the perfectly reasonable value for the film thickness, $t=134$A.

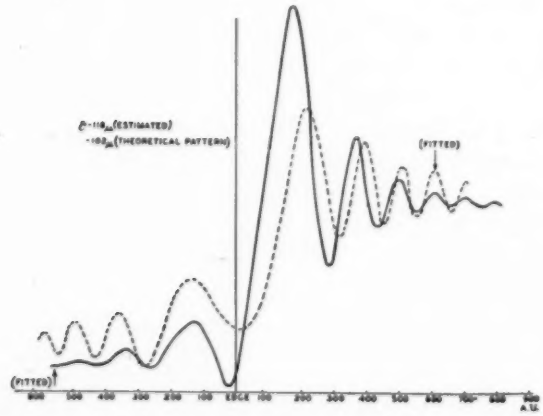


FIG. 13. Comparison of microphotometer trace and Fresnel pattern for edge of screen transmitting a quarter of incident radiation with a quarter cycle phase delay (virtual pattern).

Figure 12 shows a smoothed microphotometer trace of the contour pattern for a deviation, estimated from the lens currents used, equal to 173 microns (lens too strong). In addition, the theoretical curve for $p=0.5$ and $\Delta/\lambda=\frac{1}{4}$ has been plotted as a broken line, adjusting the value of l ($=153$ microns) and the position of the edge so that the maxima $n=3$ on the bright side and the minima $n=3$ on the shadow side of the observed and theoretical curves, respectively, are brought to coincidence. It should be stressed that these curves are comparable only with regard to the positions of the maxima and minima, not to their relative intensities. This is so for the following reasons:

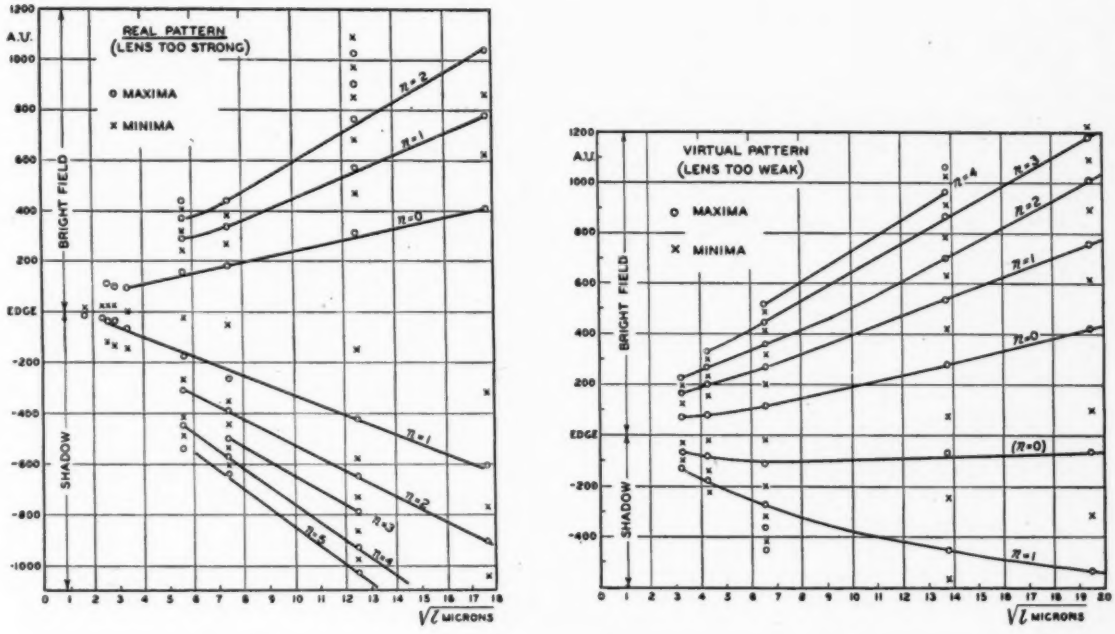
1. The theory here presented does not account for the greater intensity of the shadow portion of the edge wave as compared with the bright-field portion; this appears as a peculiarity of the Fresnel diffraction of electrons by material screens.

2. The finite illuminating aperture employed must lead to a more rapid flattening out of the maxima and minima with increasing order number in the observed curves.

3. The photometer traces shown are simple density plots, no attempt having been made to correct for the nonlinear relation between exposure and density of the photographic material employed.

Figure 13 compares the photometer trace obtained for a virtual pattern (lens too weak; estimated deviation from focus, -118 microns) with the theoretical curve for $p=0.5$, $\Delta/\lambda=\frac{1}{4}$. The fitting points were in this case the maximum $n=3$ on the bright side and the minimum $n=2$ in the shadow, the theoretical value of l required to obtain the fit being -102 micron.

For both the real and the virtual patterns the correspondence in the positions of the maxima and minima of the observed and theoretical curves, as well as the agreement between the



A. B.

FIG. 14. Variation of distance of maxima and minima from edge with square root of estimated distance between edge and focal plane. A. Real pattern. B. Virtual pattern.

estimated and the theoretical values of the deviations from focus, are close enough to render the interpretation of the observed contour fringes as Fresnel diffraction fringes quite definite.

V. BEHAVIOR OF THE CONTOURS NEAR FOCUS

If the positions of the maxima and minima relative to the shadow of the edge are plotted against the square root of the separation between the plane of the edge and the plane in which the pattern is observed, these plotted points should fall along straight lines. In practice, deviations are observed, which must in part be ascribed to inaccurate estimates of the distance between the plane of the edge and the plane of focus, as well as to the uncertainty in establishing the position of the geometrical shadow of the edge in the image (Fig. 14). For too large currents the first maximum ($n=1$) and the first minimum ($n=0$) in the shadow can be followed down to the smallest deviations from focus (Fig. 15). Their separation can conveniently be employed as a measure of deviation from focus. Figure 15 shows that these separations are generally slightly larger than the theoretical value for a normal Fresnel pattern, even with a phase delay $\Delta/\lambda = \frac{1}{4}$. For too small currents the maximum $n=0$ in the bright field and the "extra" minimum ($n=-1$) in the shadow are seen farthest. The separation of these two points (Fig. 16) also increases systematically with the deviation from focus, though only relatively slowly.

VI. IMAGE ASYMMETRIES AND THEIR ORIGIN

1. Asymmetries in Electron Micrographs

Contours of the type described above are observed quite generally in electron micrographs; in particular, the character of the fringes nearest the edge is only slightly influenced by the aperture of illumination within the range commonly employed in electron microscopy: The outlying fringes melt into the background as the aperture is increased. Micrographs obtained with an objective lens having some degree of asymmetry will show differences in the fringe patterns surrounding differently oriented edges. These differences are accentuated when the object is focused very carefully (Fig. 1). It will then occur that one edge of a particular object will exhibit the

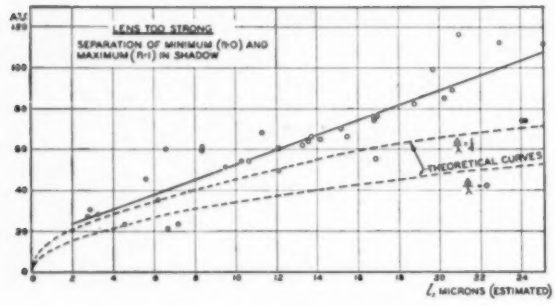


FIG. 15. Separation of first maximum and first minimum in the shadow as function of focal deviation (real pattern —lens too strong).

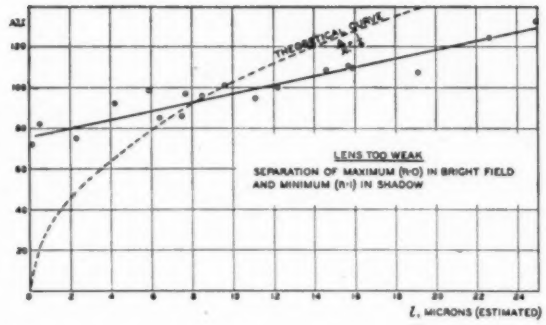


FIG. 16. Separation of first maximum on bright side and first minimum in shadow as function of focal deviation (virtual pattern—lens too weak).

characteristic virtual pattern, corresponding to too low focusing current, while an edge at right angles thereto will appear definitely overfocused. This indicates that the imaging pencils are astigmatic. The degree of astigmatism, i.e., the maximum deviation in focus for two mutually perpendicular directions, can be determined with the aid of the curves in Figs. 15 and 16. The deviation in focus for Fig. 1 may thus be estimated to be 4 microns. Two factors may give rise to asymmetries of the type described: A deviation from axial symmetry of the lens field and a misalignment of the objective.

2. Effect of Pole-Piece Ellipticity

Deviations from axial symmetry in the field of a magnetic pole-piece lens may be caused either by imperfections in machining or defects of the magnetic material in the form of inhomogeneity of susceptibility and the presence of blow holes or occlusions. Of these several factors an ellipticity of the pole-piece openings, introduced by imperfect machining, most readily admits of analytical treatment. On the basis of such a treatment it is possible to establish manufacturing tolerances in harmony with a prescribed maximum amount of astigmatism in the image.

The effect on the ray paths of deforming a permanent-magnet ring producing the lens field into an ellipse is derived in Appendix I. The permanent-magnet ring, for which the axial field distribution is shown in Fig. 17, was chosen in preference to the pole-piece lenses usually employed for reasons of ease of calculation. There can be little doubt that the qualitative conclusions—and the quantitative conclusions as far as order of magnitude is concerned—are directly applicable to the types of lenses normally used as electron microscope objectives.

The results of this study, for three objectives of the same focal length, but of varying width of the field distribution (different values of the

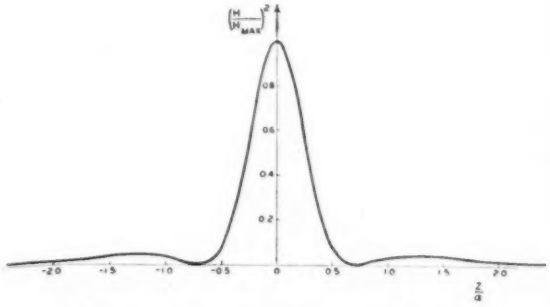


FIG. 17. Axial field distribution of permanent-magnet ring.

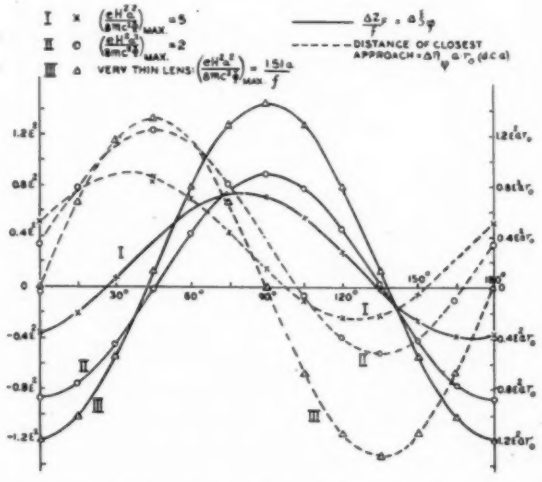


FIG. 18. Axial and lateral deviation of point of closest approach to axis from focal point of undeformed lens as function of azimuth of ray parallel to axis in image space.

TABLE I.

Len	$\left(\frac{eH^2a^2}{8mc^2\Phi}\right)_{max}$	f	a	z_F	H_{max}	χ_a	$\left(\frac{\Delta z_F}{f}\right)_{max}-\left(\frac{\Delta z_F}{f}\right)_{min}$	$(d.c.a.)_{max}$	$\Delta\varphi$
I	5	0.3	0.435	0.162	7530	−47° 33′	$1.14\epsilon^2$	$1.57\epsilon^2$	10° 7′
II	2	0.3	0.265	0.235	8050	−40° 50′	$1.76\epsilon^2$	$0.79\epsilon^2$	38′
III	0	0.3	0	0.3		0	$2.64\epsilon^2$	0	0

H = magnetic field
Φ = accelerating voltage of electrons
f = focal length of lens
a = radius of permanent-magnet ring
z_F = distance of focal point from center of lens
Δz_F = longitudinal separation of point of closest approach to axis for deformed lens from focal point of undeformed lens
$d.c.a.$ = distance of closest approach to axis for aperture angle $\alpha = 5 \cdot 10^{-3}$ radian
χ_a = total image rotation
$\Delta\varphi$ = difference in azimuth of parallel ray in image space leading to maximum value of Δz_F (weakest lens action) and of the major axis of the ellipse
ϵ = eccentricity of the ellipse: $\epsilon^2 \cong 2(a-b)/a$, where a is the major and b is the minor axis of the ellipse

radius of the magnetic ring), are summarized in Fig. 18 and Table I. A ray which is parallel to the axis in image space will intersect the axis in object space for only four specific values of the azimuthal angle φ, measured relative to the meridional plane containing the long axis of the ellipse. For all other values there will exist a distance of closest approach ($d.c.a.$) to the axis, at a point a distance Δz_F from the focal point of the undeformed lens (Fig. 19). The ratio of Δz_F to the focal length f and the distance of closest approach are plotted as functions of the azimuthal angle in image space, φ, in Fig. 18. The results shown lead to the following conclusions:

1. An ellipticity of the pole-piece bore results in an astigmatism of the imaging pencils. The distance between the two points of sharpest focus (the focal lines) is proportional to the percentage difference between the two axes of the ellipse. For a focal length of 3 mm a difference in the two axes equal to 0.01 percent ($\epsilon^2 = 2 \cdot 10^{-4}$) leads to a separation of the two focal lines of the order of one micron.

2. The ratio of the distance between focal lines and the focal length decreases as the ratio of lens thickness to focal length is increased. This may be ascribed to the fact that, as an imaging ray rotates in the field, the ray is acted upon both by portions of the field which are strengthened and such as are weakened by the deformation. The amount of the rotation is greater for a thick lens than for a thin lens.

3. For weak or short lenses edges in the image which, for proper adjustment of focus, may be made sharp, are parallel to the two axes of the ellipse (edges which appear sharp at the higher lens current are parallel to the minor axis). For strong or thick lenses the two orientations corresponding to the sharpest possible focus are shifted by a small angle with respect to the axes of the ellipse.

4. For weak or short lenses the astigmatic pencil has two focal lines, so that the apparent sharpness of object edges oriented parallel to these, duly refocused, is not reduced by the deformation. For stronger lenses the focal lines become flat ellipses. Under these circumstances—i.e., the presence of "anisotropic astigmatism"—the deformation results in some loss of sharpness throughout the image. The focal ellipses are sufficiently flat, however, that the diffusion of the edges parallel to their axes becomes negligible if the asymmetry in the image is not apparent.

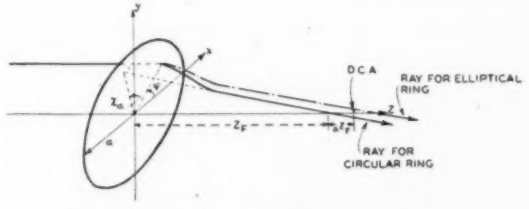


FIG. 19. Geometrical parameters of rays.

3. Effect of Misalignment of the Objective

Asymmetries in the image may also be introduced by a horizontal displacement of the objective relative to the projector or a tilt of the objective lens relative to the instrument axis. If S_4 is the aberration coefficient of astigmatism[9] for the objective, M its magnification, and f its focal length, the displacement of the two focal lines referred to the object becomes

$$\Delta z_{F\,max} - \Delta z_{F\,min} = 2(S_4/M) \cdot f \cdot (d/M)^2 \quad (2)$$

for a displacement d of the objective relative to the projector and

$$\Delta z_{F\,max} - \Delta z_{F\,min} = 2(S_4'/M) \cdot f^3 \cdot \delta^2 \quad (3)$$

for a tilt angle δ of the objective. The coefficient S_4 differs from the coefficient S_4', since the first applies to a position of the effective limiting

[9] $S_4 r_0^2 r_a$ is the radius of the circle of confusion owing to astigmatism which forms the image of an object point a distance r_0 from the axis, imaged by a pencil whose cross-section radius in the aperture plane is r_a. See V. K. Zworykin, G. A. Morton, E. G. Ramberg, J. Hillier, and A. W. Vance, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945), pp. 551, 562, and 636.

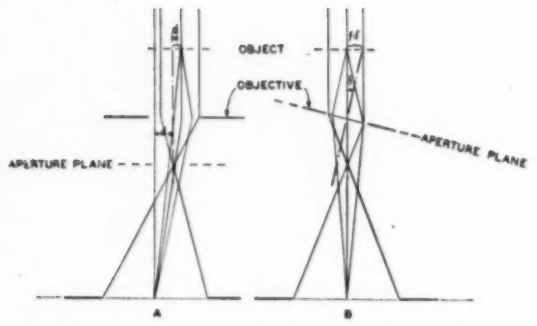


FIG. 20. Passage of imaging pencil through displaced objective (A) and through tilted objective (B).

aperture at the rear focal plane of the objective, the second to one at its nodal point (Fig. 20). To take only one example, the values of S_4/M and S_4'/M for a magnetic objective formed by two unsaturated equidiameter cylindrical pole pieces with a small separating gap are for $\Phi = 50$ kv, $f = 0.3$ cm, $H_{max} = 6270$ gauss, radius $a = 0.575$ cm, which corresponds to a very thick, strong lens,

$$S_4/M = 16 \text{ cm}^{-2}; \quad S_4'/M = 1.5 \text{ cm}^{-2}.$$

For a permissible difference in focus of 1 micron (10^{-4} cm), the tolerance in the displacement d and the tilt δ become

$$d = \pm 0.3 \text{ cm}; \quad \delta = \pm 0.035 \text{ radian} = \pm 2.0°.$$

Both tolerances are seen to be quite large; the normal alignment procedure should render the effects of objective displacement negligible.

VII. METHOD OF ELIMINATING ASYMMETRIES

When it was realized that the asymmetries in the magnetic field of the electron microscope objective were mainly responsible for the limitations on the resolving power it was assumed at first that these asymmetries were the result of mechanical inaccuracies—elliptical openings and misalignment of the openings of the pole pieces. To overcome these, a lapping technique was devised which insured the required precision of mechanical symmetry and alignment. As will be shown later, mechanical inaccuracies were actually only partially responsible for the asymmetries observed. However, a remarkable improvement[1] was obtained in most cases by the use of the lapping technique and for that reason a brief description is included.

The pole pieces used in these tests are standard objective lens pole pieces for the RCA EMU electron microscope—actually rejects from the commercial production—used *without* an objective aperture. They consist of two soft-iron sections bearing the active pole faces and openings and part of the magnetic circuit of the objective coil unit. The iron sections are threaded into a brass spacer which maintains the proper spacing and alignment through the provision of appropriate pilot surfaces. The cross section of these pole pieces is shown schematically in Fig. 21.

A conical brass lap with a 1°–2° taper was used for lapping the openings. The pole faces were lapped on a plate-glass surface, a large brass jig being used to hold the axis of the lower section accurately perpendicular to the plate. The two openings were lapped simultaneously with the pole pieces assembled in order to obtain accurate axial alignment as well as symmetry. It was necessary, of course, to disassemble the pole pieces in order to gain access to the pole faces for lapping or for cleaning. In order to avoid as much as possible differences in alignment of the two sections occurring as a result of the disassembling, fine scratches were made on the outside of the pole-piece assembly across the boundaries between the pole-piece sections and the brass spacer. Each time the pole pieces were assembled, great care was taken to clean the threads and pilot surfaces and to align the scratches accurately.

The procedure followed was to lap, alternately, lens openings with the lens assembled and faces, individually, with the same grade of grinding compound. A wet paste of 400-mesh carborundum was used for the initial lapping which was continued until all machining marks and other defects disappeared. Lapping was then continued with wet pastes of 600-mesh carborundum and 900-mesh alundum, lapping with the last two grades of grinding compound being continued until the effects of the preceding grade were removed and until there appeared to be a perfect fit between the brass lap and the pole-piece openings for all angular positions of the lap. Considerable care was taken to avoid exerting too much pressure with the lap as this was found to damage the critical edge between the pole-piece openings and the faces. When the grade of

grinding compound was changed the pole pieces were very carefully washed to remove all traces of the coarser grade. After the lapping was completed with the 900-mesh alundum and if no imperfections were visible with the 10× eye lens, particularly in the regions near the critical edges, those regions of the pole-piece openings and faces were polished using a wet paste of jeweler's rouge on a round toothpick.

Pole pieces which were treated in this way invariably showed considerable improvement in performance and rather consistently gave limiting resolving powers in the range of 20–30A (Fig. 22). It was soon realized, however, that the asymmetries were seldom completely removed by this technique. Moreover, if considerable asymmetry existed after the first lapping, it was usually found that it persisted upon further lapping. In cases where the asymmetry was found to increase with further lapping, it could generally be shown to arise from defects in the magnetic material of the pole piece, in particular, blow holes or occlusions which occurred in the neighborhood of the critical edge and which would eventually be uncovered. These observations indicated rather conclusively that the asymmetries remaining after lapping were caused by various forms of imperfections in the magnetic properties of the pole-piece material.

Considerable effort was expended in an attempt to obtain ideal magnetic materials. Up to the present none has been found. In fact, the outcome of this work pointed to the empirical correction of the lens asymmetries as being more likely of success and less tedious in execution.

Asymmetries of the magnitude encountered in the uncorrected objective may be considered as a superposition of a cylindrical magnetic lens on an accurately axially symmetrical magnetic lens. The practical method of correction involves, therefore, the production of an additional superimposed cylindrical field which is of the proper strength and at right angles to the existing cylindrical component. There are many possible ways of accomplishing this. The device selected is illustrated in Fig. 23. It consists simply of eight soft-iron rods 0.070 inch in diameter threaded into the wall of the brass spacer (72 threads to the inch), perpendicular to the axis of the lens in the plane of the pole-piece gap and 45° apart. Figure 24 is a photograph of a pole piece with the compensating screws in place.

The procedure of compensation is as follows:

1. A focal sequence (Series 1) is obtained under the conditions to be described below with the rods adjusted so that they are approximately equidistant from the axis of the lens. The orientation and magnitude of the asymmetry is noted on the appropriate exposure (Fig. 25).

2. Any opposite two of the rods are moved as far as possible toward the axis and another focal sequence (Series 2) is obtained under conditions identical to those of Series 1. This operation puts in an extreme degree of asymmetry and permits a correlation to be made between the direction of the asymmetry in the image and the direction of adjustment of the rods in the lens (Fig. 26).

3. The pair of rods which must be moved toward the axis in order to correct for the asymmetry existing in Series 1 is then determined by a comparison of Series 1

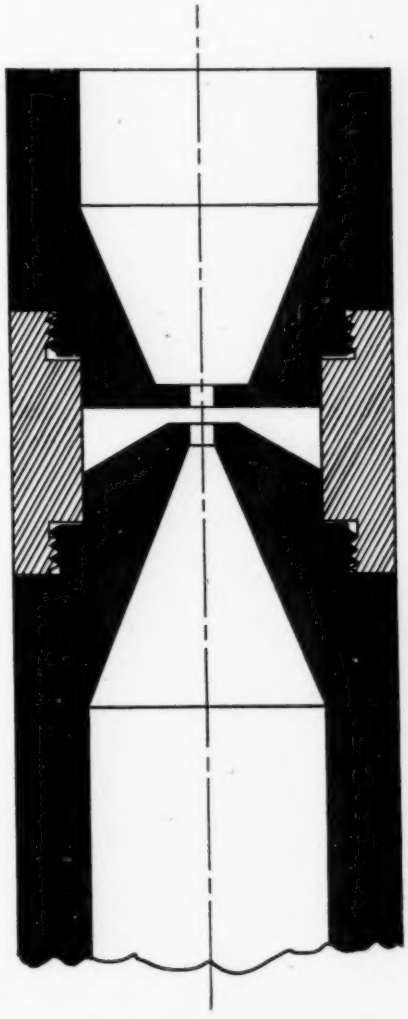


FIG. 21. Cross section of objective pole pieces.

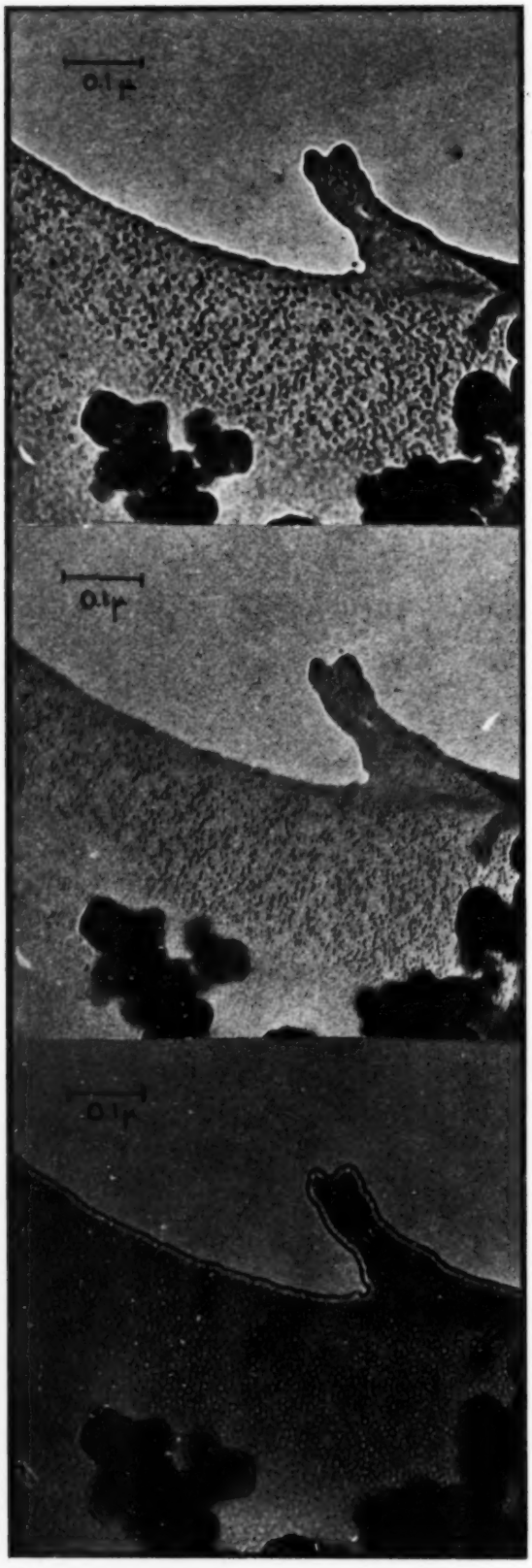


FIG. 22. A focal series obtained with a lapped lens, showing fairly complete elimination of asymmetries. The specimen is a very light evaporated film of gold on a carbon black in collodion mount. This series shows the rapid increase in the diameter of the images of small particles which occurs as the instrument is put out of focus.

and 2. The compensation is of course perpendicular to the pair of rods which would increase the asymmetry observed in Series 1.

4. Starting with the rods equidistant from the axis the two selected for the first correction are moved toward the axis and a third focal sequence (Series 3) is obtained. The direction and degree of asymmetry remaining is determined from the proper image (Fig. 27).

5. If the asymmetry is found to be decreased but has not changed direction the same two rods are moved closer to the axis and another focal sequence (Series 4) is obtained. As shown in Appendix II, the maximum correction attainable by moving any one pair of rods toward the axis is reached when the direction of the asymmetry in the image is rotated through an angle of 45° (Fig. 28).

6. With the rods used in the initial correction set at the point where the direction of the asymmetry has been rotated by an angle in the range of 30° to 45° from its original direction a new direction of correction is determined by comparison of the last series and Series 2. With the rods in the newly selected direction advanced slightly toward the axis another focal series is obtained (Series 5—Fig. 29).

7. Correction is continued in the newly selected direction until the direction of asymmetry again changes.

8. This systematic compensation of the asymmetries is continued until no detectable asymmetry exists in the images (Fig. 30).

In actual practice it is found that the asymmetry decreases very rapidly and a third direction of compensation is seldom necessary.

VIII. INSTRUMENT ADJUSTMENT

In discussing the correction of a lens by the above method no mention has been made of the adjustment of the other components of the electron microscope employed. It is obvious that all the images used in the course of the correcting procedure must be obtained under identical electron-optical conditions. It has been found, furthermore, that the compensation obtained by the above technique is accurate for only *one* adjustment of the instrument. As would be expected, the method described corrects not only the asymmetries inherent in the lens, but also those introduced by stray fields from the other components of the instrument. The various phenomena encountered and the adjustments

which must be carried out if the procedure is to give consistent results are listed below:

1. Specimen.—A suitable specimen must be used. Small holes in the collodion membrane or spherical carbon-black particles mounted without a supporting membrane have been satisfactory in this laboratory. ***Crystalline materials are to be avoided*** because of the artificial asymmetries sometimes introduced by different interactions of the electron beam with different faces of the crystals.

2. Illumination.—The angular aperture of the illumination should be kept as small as possible in order to obtain the sharpest possible definition in the contour fringes. In the present work the angular aperture of illumination was, in general, below 10^{-4} radian, though it might have been as much as an order of magnitude larger without seriously limiting the method. The illumination must arise from a single source.

3. Lens Position.—The objective pole-piece assembly must be replaced in exactly the same position in the objective lens coil structure each time it is put back after making an adjustment. This concerns only the angular orientation since the axial position is usually fixed by the structure of the lens.

4. Specimen Position.—It is obvious that an empirical method of compensating for lens field asymmetries arising as a result of imperfections in the magnetic properties of the material will be precise for only one condition of magnetization. This means that the specimen position and the accelerating potential of the electrons must be accurately reproduced both during the compensation of the lens and during subsequent high resolution work.

5. Magnification.—The preceding condition insures that the objective magnification be constant during its compensation and use. Because of the stray fields produced by the projection lens, and the dependence of the alignment on the power of the projection lens, it was found necessary to keep its magnification constant, preferably at its maximum value. The total instrumental magnification used in the present tests was 16,000× for the objective shown in Fig. 23 and 25,500× for the standard pole pieces used in subsequent tests.

6. Excitation of the Objective Lens.—External compensation in a magnetic lens field of asymmetries which are the result of irregularities in the susceptibility of the magnetic material requires, as already mentioned, that the same condition of magnetization be maintained during the compensation and use of the lens. The purpose of this particular section is to point out that the same condition of magnetization cannot be achieved unless the lens current is turned on in such a way as to insure that every region of the material of the pole pieces is at the same point on its hysteresis loop every time the lens is operated, either during the compensation procedure or in use. The procedure which was arbitrarily selected for the present work and which was found to be satisfactory is as follows: The image is focused visually and the objective current turned off. It is left off for about ten seconds to allow the circuits

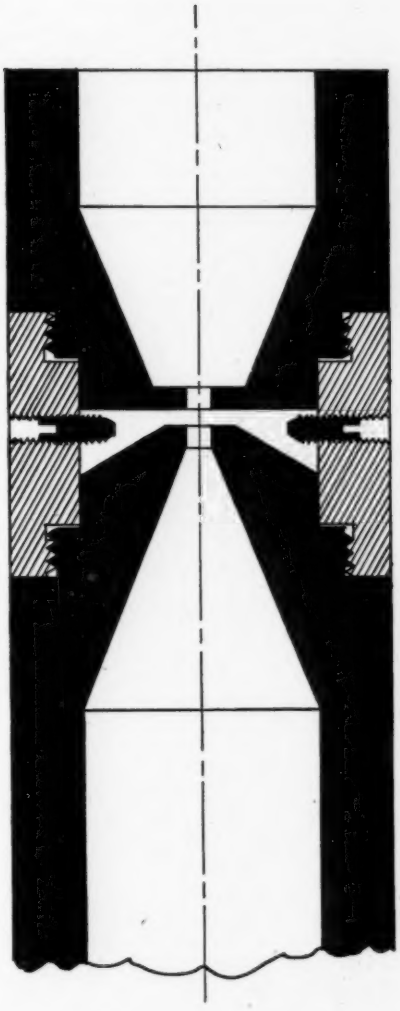


FIG. 23. Cross section of objective pole pieces provided with compensating screws.

FIG. 24. A photograph of the objective pole piece with the top removed to show the compensating screws.

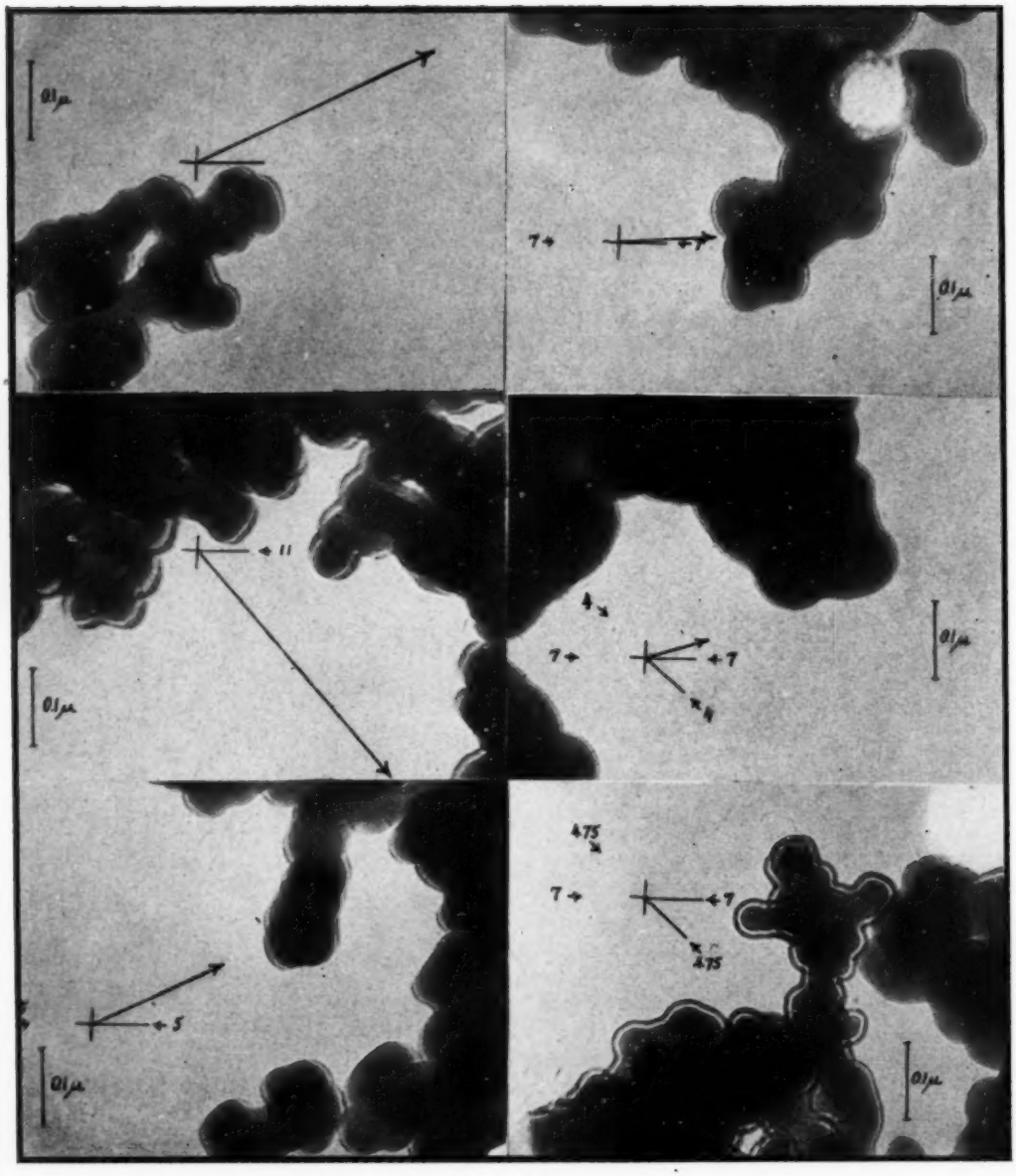


FIG. 25

FIG. 28

FIG. 26

FIG. 29

FIG. 27

FIG. 30

FIG. 25. Focused image obtained with an uncompensated objective.

FIG. 26. Focused image obtained with one pair of correcting screws turned in as far as possible (11 turns). This puts in a large asymmetry which can be correlated with the compensating direction.

FIG. 27. First stage of compensation. The appropriate pair of screws has been selected and has been moved in five turns. The asymmetry is in the same direction as in the initial test (Fig. 25) but of smaller magnitude.

FIG. 28. Illustrates the maximum obtainable compensation with the one pair of screws (7 turns). The magnitude of the asymmetry is greatly decreased and its direction has changed.

FIG. 29. Shows compensation in the selected second direction (4 turns).

FIG. 30. Image obtained with fully compensated objective. The particles in the lower right hand corner are at exact focus. Other particles are either considerably above or below focal plane. Differences in level are emphasized in the compensated objective.

to become stabilized and then turned on again. The image is then observed and refocused. This procedure is repeated until the image comes into exact focus without further adjustment every time the lens current is turned on. The above technique was found to obliterate the previous magnetization history of the pole pieces and to produce consistent results as judged by the observed asymmetries.*

* *Note added in proof:* It has been found that the standard objective lens controls were not sufficiently fine for the most critical work. To overcome this a vernier has been

7. *Instrument Alignment.*—As a result of the present investigations the ideas on instrument alignment previously accepted in this laboratory underwent considerable revision. It was found that the quality of the images obtained with a reasonably symmetrical and clean objective lens was insensitive to alignment if considered from the point of view of astigmatism, but was quite sensitive considered from the point of view of chromatic change in magnification and rotation. This is in agreement with theory. The variations in the accelerating potential were measured and found to consist, primarily, of a 3-volt (peak to peak) 60-cycle ripple, which amounts to 0.006 percent of the total voltage and is well within the permitted variations for the attainment of a resolving power of 10A. This tolerance is valid, however, only if the "voltage center" (that is, the point of the image which remains stationary with change in accelerating potential) coincides with the center of the image field. In the present work it was found that, as a result of asymmetries associated with the projection lens, the magnetic and voltage centers were widely separated when the instrument was aligned on the basis of image rotation with objective current change and that the chromatic variation in magnification appeared as a 40A image shift.**

It is obvious from these observations that the alignment should be carried out on the basis of voltage rotation. The alignment procedure actually employed is as follows:

a. Center objective over projector mechanically.
b. Turn the accelerating voltage off and on very rapidly at the lowest projector setting and with the condenser adjusted to maximum power at which an image can still be seen, observing the effect on the image.
c. Translate condenser until center of voltage rotation or expansion is in center of field.
d. Reduce condenser current and translate gun until maximum illumination of field is obtained.
e. Alternately repeat steps c and d until voltage center remains in field for all condenser currents.
f. Raise magnification to maximum value and bring voltage center into field by translating objective.
g. Check e.

When the compensation of the objective was carried out according to the routine outlined above and the instrument adjustments, also listed above, were checked every time the objective lens was inserted for measurement it was found that the compensation was accomplished in accord with theory and that any setting of the lens-correcting screws gave reproducible results. It was found further that under the same condition a fully compensated lens was stable over reasonably long periods. One standard compensated lens maintained its symmetry over a two-week period of constant operation by one operator. In the succeeding two-week period, during which time the instrument was operated by several operators, the same lens became very slightly asymmetrical. However, since accurate records of the results of compensating adjustments in the original correction of the lens had been kept, it was a relatively simple matter to re-attain complete symmetry in the lens.†

IX. ACCIDENTAL INSTRUMENTAL DEFECTS

Throughout the preceding discussion it has been assumed that the electron microscope being used in these tests performed perfectly within the limitations inherent in its design. Electron microscopists are aware, however, that such is seldom the case in practice and that the attainment of consistently high resolving power is only possible if the operator is continuously on the alert for indications of the presence of any of a large number of possible accidental defects. This situation is accentuated when resolving powers better than 20A are desired. In the following, a number of the more common accidental defects which are particularly important in the high resolving power range will be described and some methods for their detection and elimination given.

1. External Magnetic Fields

Commercial electron microscopes are normally carefully shielded for stray alternating magnetic fields, so that the image defects due to such fields are reduced below the 20A level. However, in the attainment of high resolving powers, particularly in the case of unusually bad locations, such stray fields still are an important limiting factor. In this laboratory an image shift of 20A (using an objective lens with a 2.5-mm focal length) was traced to this source.

The conventional electron microscope is most

added for which the total variation is 1/10 of that of what was previously the finest control.

** *Note added in proof:* The separation between the magnetic and voltage centers has now been eliminated by the introduction of a soft iron tube immediately above the projection lens pole-piece opening and extending to the top of the coil form. This tube was mounted coaxially with the projection lens system and magnetically isolated from it.

† *Note added in proof:* Since the second correction described above was made, this lens has remained accurately symmetrical over a period of six months during which time the pole pieces have been removed from the instrument and replaced well over 100 times.

sensitive to the effects of alternating magnetic fields in the region of the instrument immediately below the objective lens. Thus, the direction and absolute magnitude (that is, *not* referred to the scale of the specimen plane) of the image shifts resulting from alternating magnetic fields must be constant for all images taken with constant projection lens power and must be independent of the objective lens strength and alignment. These properties are sufficient for the identification of image defects due to alternating magnetic fields.

A single additional shield of small diameter ($\frac{1}{2}$ in.) mu-metal tubing placed along the axis of the instrument between the objective and the projection lenses was sufficient to eliminate these defects.

2. Specimen Vibration

In the present work some image shifts which were fairly consistent in the 10–20A range and erratic in the 20–40A range were finally traced to vibrations of the specimen holder. Since the resolving power of the instrument for visual observation through the final image telescope[1] at 250,000× is approximately 30A only the larger transitory vibrations could be seen. These were correlated with the boiling of the oil in the diffusion pump and the operation of the shutter.

This vibration, which was extremely troublesome in critical work, was finally eliminated by the simple expedient of lengthening the specimen holder until it just touched the upper face of the objective pole pieces. Unfortunately, under these conditions the stage movement controls had considerable backlash and were extremely difficult to manipulate. Thus, while the technique served for the present work, it is not considered a satisfactory solution of the problem.

3. Specimen Stage Drift

In the course of the present work it was found that the specimen stage did not remain motionless after being adjusted but tended to drift by as much as 30A during a 30-second exposure. The effect could be partially eliminated by carefully relieving mechanical strains in the stage mechanism by appropriately adjusting the controls and observing the results at high magnification. The above-mentioned technique of steadying the specimen, with its attendant difficulties, was most successful in eliminating specimen stage drift.

4. Specimen Drift (Thermal)

When the angular aperture of the illumination is varied from its maximum value for visual observation to a considerably lower predetermined value for photographic recording the total electron current reaching the specimen remains constant, but its distribution over the specimen changes considerably. This is particularly true in the instrument used for the present work. Since, at maximum angular aperture, the area of irradiation is only 10 microns in diameter, the wire supporting screen is seldom bombarded. On the other hand, at an angular aperture of $5 \cdot 10^{-5}$ radian, at which the micrographs are recorded, the area of irradiation is 330 microns in diameter and, hence, includes several meshes of the wire screen. It is obvious that when the change is made a new condition of thermal equilibrium in the specimen screen must be established and that this may involve a change in position of parts of the specimen screen owing to thermal expansion. It has been found that image shifts due to such thermal movements are very common unless special precautions are taken to avoid them. Woven screens are particularly bad in this respect because individual wires can move independently of the rest. Electrolytically produced copper screens give the best results at the present time, though even these are not always completely satisfactory. The screens must be fairly heavy, which means, of course, that the open area is small (20 percent). Special care must be taken with the clamping of a screen in the specimen holder to insure that it is held tightly around its entire circumference.

5. Charging of the Specimen

In accord with observations made with the probe diffraction camera it was found that the area of the specimen irradiated became highly charged *if it did not include a grounded conductor*. While the charging of the specimen does not affect the image quality as far as can be determined by visual observations it does give rise to

erratic image shifts, making it impossible to attain high resolution micrographs. These image shifts can be completely eliminated by spreading the beam until one of the screen wires is included in the area of irradiation. Under these conditions the entire area of irradiation is held at the same potential as the screen. The mechanism of this discharging, which appears to be the same as encountered in the use of electron "sprays" in electron diffraction, is not yet understood.

It was also found necessary to take special precautions to clean the specimen screens before use. Large particles of lint or other insulating materials outside the area of irradiation often collect sufficient charge to deflect the beam and produce image shifts.

6. Contamination of Specimen Holder

Insulating particles and remnants of previous specimens clinging to the edges of the specimen cap are also a common source of image shifts. The existence of such particles can usually be detected by the occurrence of changes in image position with changes in intensity. In this laboratory the specimen cap and the surface of the hole through the specimen holder are kept polished by frequent cleaning with a very fine grade of steel wool wrapped around a round toothpick. The specimen cap is examined with a 10× eye lens for such contamination before inserting the specimen every time high resolution images are desired. Obviously, care must be taken to avoid leaving any loose threads of the steel wool on the holder.

7. Contamination of Objective Lens

Insulating particles clinging to the surfaces of the pole-piece openings introduce both image shifts and image asymmetries and are a continual source of trouble. As in the preceding case the existence of such particles makes itself apparent through a change in image position with change in intensity. The particles can be removed satisfactorily with lint-free lens tissue[10] on a toothpick, a 10× eye lens being used to observe the operation.

[10] Ross-Adams Lens Tissue distributed by Clay-Adams Company, Inc., New York, New York.

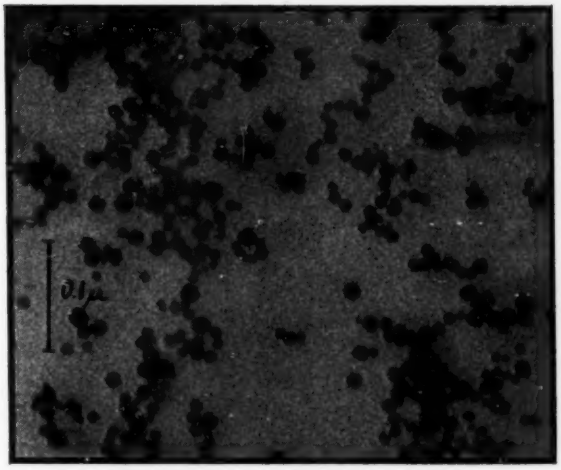


FIG. 31. High magnification image of colloidal gold taken with compensated objective ($f = 3.8$ mm).

8. Change of Focus with Change of Illumination

In the early critical work the change of focus which occurs when the intensity and angular aperture of illumination is changed gave considerable difficulty for the narrow range focal series used. The use of the high intensity gun and telescopic viewing device making it possible to focus the instrument and make the photographic exposure without changing the adjustment of the illumination eliminated the difficulty.

9. General Remarks

As the reader will have noticed, all of the above defects result in image shifts. If several are present simultaneously and they are observed only in the recorded image it is difficult, if not impossible, to determine their cause. In the present work it has been possible to sort out the various effects and to determine their cause individually through the use of telescopic observation of the image at high magnifications. This permits transitory image motions to be correlated with other incidental phenomena in the operation of the instrument. It is hoped that the presentation of a description of these defects in the above paragraphs will enable other electron microscopists to systematically eliminate them in their own particular instruments, which may not be provided with telescopic observation of the image. It is realized that the above list is not complete and, in particular, does not include

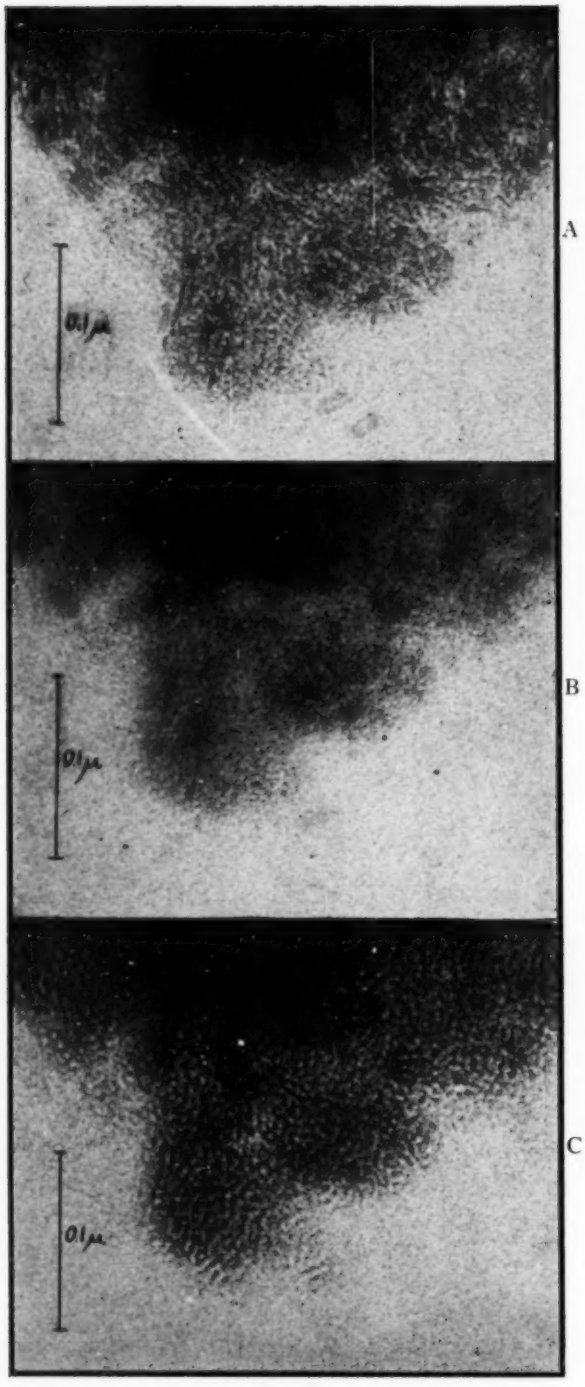


FIG. 32. High magnification focal series of alumina gel taken with compensated objective ($f=2.4$ mm). A. Underfocused image; B. Focused image; C. Overfocused image.

a description of the methods of identifying specific defects in the electrical supplies. The latter can be easily deduced, however, from a study of the normal operation of the instrument.

X. OBJECTIVE APERTURE

Throughout the development of the electron microscope there have been many differences of opinion with regard to the necessity and value of a physical limiting aperture in the objective. The present work, which was carried out entirely without an aperture, has shown conclusively that such an aperture is not necessary for the demonstration of resolving powers at the 10A level. On the other hand, it has become apparent that lack of contrast at exact focus is at present the major limiting factor in the resolution of extremely fine structures of organic materials. A few attempts, which were made in the course of the present work, to establish the value of a limiting aperture in this particular problem were not conclusive.

XI. CONCLUSION

It is demonstrated that the Fresnel diffraction fringes which are present in extra-focal images

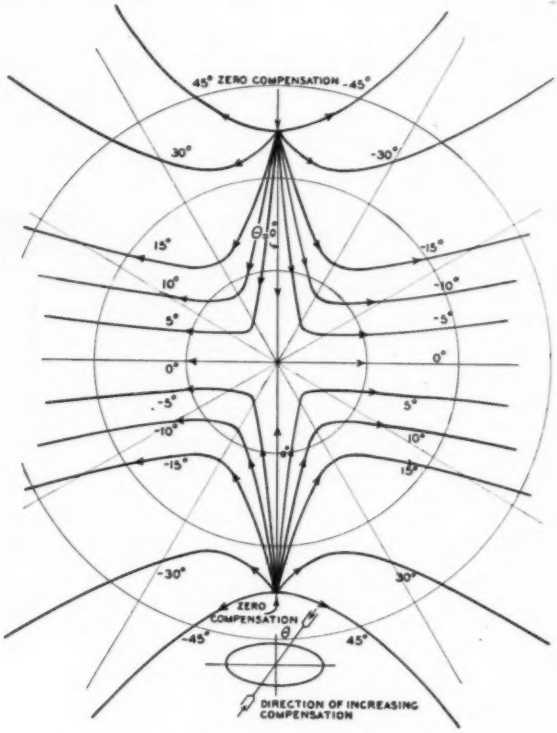


FIG. 33. Variation of orientation and magnitude of residual asymmetry with magnitude of compensating field for various angles between the principal axes of the original field and the compensating field.

obtained with small angular aperture of illumination provide a sensitive criterion of the degree of symmetry possessed by an electron microscope objective. It is shown further that the use of such a criterion makes the correction of asymmetries a relatively straightforward and simple matter. The image quality obtained with a corrected lens (Figs. 31 and 32) has been consistently and obviously higher than that obtained with the best uncorrected lenses.[1] While it has not been possible to make a precise measurement of the best resolving power achieved it is believed to lie in the range of values predicted theoretically for the lenses used.

The observed contour fringes are found to correspond to the Fresnel diffraction pattern for the edge of a partly transparent screen which introduces a phase delay in the coherently transmitted wave. The phase delay, arising from the inner potential of the screen, and the empirically observed greater intensity of the edge wave in the shadow, account for the observed dissimilarity of the contours for underfocusing and overfocusing.

Since the technique of obtaining symmetry in the properties of the image corrects asymmetries owing to external causes as well as those produced by defects in the lenses and since this compensation is not independent of instrument adjustment it appears, at present, that the attainment of high resolving power in a particular instrument requires the individual attention of the electron microscopist responsible for its operation. A number of practical suggestions have been included in the above discussion in the hope that they will assist the interested electron microscopist in the recognition and elimination of some of the more common instrumental difficulties which he may encounter.

APPENDIX I

Field Variation and Electron Paths for Elliptical Permanent Magnet Ring

If q is the magnetic moment per unit length of the magnetic ring, a (along the x axis) is the major axis, and $b=a(1-\epsilon^2)^{\frac{1}{2}}$ (along the y axis) is the minor axis, the scalar potential variation in the plane containing the major axis is

$$\psi(\xi, 0, \zeta)=\frac{q\zeta}{a}\int_{-1}^{1}\frac{(1-\epsilon^2\xi_0^2)^{\frac{1}{2}}d\xi_0}{(1-\xi_0^2)^{\frac{1}{2}}[(\xi_0-\xi)^2+(1-\epsilon^2)(1-\xi_0^2)+\zeta^2]^{\frac{1}{2}}}. \tag{4}$$

Here $\zeta=z/a$, $\xi=x/a$, x and z being the coordinates of the reference point. For the meridional plane containing the minor axis,

$$\psi(0, \eta, \zeta)=\frac{q\zeta}{a}\int_{-1}^{1}\frac{(1-\epsilon^2\xi_0^2)^{\frac{1}{2}}d\xi_0}{(1-\xi_0^2)^{\frac{1}{2}}\{\xi_0^2+([(1-\epsilon^2)(1-\xi_0^2)]^{\frac{1}{2}}-\eta)^2+\zeta^2\}^{\frac{1}{2}}}, \tag{5}$$

$$\eta=y/a. \tag{6}$$

Near the axis a Taylor expansion of the potential yields

$$\psi=\Psi(\zeta)+\frac{1}{2}\left(\frac{\partial^2\psi}{\partial\xi^2}\right)(0, 0, \zeta)\xi^2+\frac{1}{2}\left(\frac{\partial^2\psi}{\partial\eta^2}\right)(0, 0, \eta)\eta^2+\cdots. \tag{7}$$

This is to be compared with the Taylor expansion for the corresponding axially symmetric system ($\epsilon=0$)

$$\psi_0=\Psi_0(\zeta)-\tfrac{1}{4}\Psi_0''(\zeta)r^2+\cdots, \tag{8}$$

where

$$\Psi_0(\zeta)=\frac{\pi q\zeta}{a(1+\zeta^2)^{\frac{1}{2}}}, \quad \Psi_0'(\zeta)=\frac{\pi q}{a}\frac{(1-2\zeta^2)}{(1+\zeta^2)^{5/2}}. \tag{9}$$

For small values of ϵ the differences of corresponding coefficients in the two expansions are simply proportional to ϵ^2:

$$\frac{\Delta\Psi'}{\Psi_0'}=\frac{1}{\Psi_0'}(\Psi'-\Psi_0')=\frac{\epsilon^2}{4}\frac{4\zeta^4-10\zeta^2+1}{(1+\zeta^2)^{7/2}}, \tag{10}$$

$$\frac{1}{\Psi_{0_{max}}'}\Delta\left(\frac{\partial^2\psi}{\partial\xi^2}\right)=\frac{1}{\Psi_{0_{max}}'}\left(\frac{\partial^2\psi}{\partial\xi^2}+\frac{\Psi_0''}{2}\right)=\frac{3\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}\left(\frac{1}{4}-\frac{27}{16}\zeta^2+\frac{\zeta^4}{4}\right), \tag{11}$$

$$\frac{1}{\Psi_{0_{max}}'}\Delta\left(\frac{\partial^2\psi}{\partial\eta^2}\right)=\frac{1}{\Psi_{0_{max}}'}\left(\frac{\partial^2\psi}{\partial\eta^2}+\frac{\Psi_0''}{2}\right)=\frac{3\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}\left(\frac{11}{4}-\frac{57}{16}\zeta^2+\frac{\zeta^4}{4}\right), \tag{12}$$

and hence

$$\frac{1}{\Psi_{0_{max}}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)=\frac{3}{4}\frac{\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}(12-21\zeta^2+2\zeta^4), \tag{13}$$

$$\frac{1}{\Psi_{0_{max}}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)=-\frac{15}{8}\frac{\epsilon^2\zeta}{(1+\zeta^2)^{9/2}}(4-3\zeta^2). \tag{14}$$

The equations of motion of the electrons in the elliptical field are, in the paraxial approximation

$$\frac{\partial^2\xi}{\partial t^2}=\frac{e}{mca}\left(\dot\eta\Psi'-\frac{\partial^2\psi}{\partial\eta^2}\dot\zeta\eta\right);\quad \ddot\eta=\frac{e}{mca}\left(\frac{\partial^2\psi}{\partial\xi^2}\dot\zeta\xi-\dot\xi\Psi'\right);\quad \frac{\partial^2\zeta}{\partial t^2}=0. \tag{15}$$

Here dots indicate differentiation with respect to the time. The equations of motion may be converted into the path equations

$$\eta''=\frac{\ddot\eta\dot\zeta-[(\partial^2\zeta/\partial t^2)\dot\eta]}{\dot\zeta^3}\cong\frac{\ddot\eta}{\dot\zeta^2}=\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left(\frac{\partial^2\psi}{\partial\xi^2}\xi-\xi'\Psi'\right), \tag{16}$$

$$\xi''=\frac{[(\partial^2\xi/\partial t^2)\dot\zeta]-[(\partial^2\zeta/\partial t^2)\dot\xi]}{\dot\zeta^3}\cong\frac{(\partial^2\xi/\partial t^2)}{\dot\zeta^2}=\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left(\eta'\Psi'-\frac{\partial^2\psi}{\partial\eta^2}\eta\right), \tag{17}$$

since $\dot\zeta=(2e\Phi/m)^{\frac{1}{2}}/a$ where Φ is the accelerating potential of the electrons. Primes indicate differentiation with respect to ζ. Equations (16) and (17) may be combined into a single equation:

$$(\xi+i\eta)''+i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left[\Psi'(\xi+i\eta)'-\left(\frac{\partial^2\psi}{\partial\xi^2}\xi+i\frac{\partial^2\psi}{\partial\eta^2}\eta\right)\right]=0 \tag{18}$$

or

$$(\xi+i\eta)''+i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left[\Psi_0'(\xi+i\eta)'+\frac{\Psi_0''}{2}(\xi+i\eta)\right]$$

$$=-i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left\{\Delta\Psi'(\xi+i\eta)'-\left[\xi\Delta\frac{\partial^2\psi}{\partial\xi^2}+i\eta\Delta\frac{\partial^2\psi}{\partial\eta^2}\right]\right\}. \tag{19}$$

Introduce now the new independent variable

$$u=(\xi+i\eta)e^{-i\chi};\quad \chi=-\int_{-\infty}^{\zeta}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_0'd\zeta. \tag{20}$$

The differential Eq. (19) now becomes:

$$u''+\frac{e\Psi_0'^2}{8mc^2\Phi}u=-i\left(\frac{e}{2mc^2\Phi}\right)^{\frac{1}{2}}\left\{\Delta\Psi'\left(u'-i\left[\frac{e}{8mc^2\Phi}\right]^{\frac{1}{2}}\Psi_0'u\right)-e^{-i\chi}\left[\xi\Delta\frac{\partial^2\psi}{\partial\xi^2}+i\eta\Delta\frac{\partial^2\psi}{\partial\eta^2}\right]\right\}=R(\zeta). \quad (21)$$

The corresponding homogeneous equation,

$$u_0''+\frac{e\Psi_0'^2}{8mc^2\Phi}u_0=0, \quad (22)$$

has the particular solution u_a with the initial conditions $u_a=0$, $u_a'=-a/f$ for $z=z_F$; for $z=-\infty$, $u_a=1$.

Consider now a ray incident parallel to the axis, unit distance from it, and in a meridional plane with the azimuth φ. In the zero approximation its coordinates are then given by

$$u=u_ae^{i\varphi}; \quad \xi=u_a\cos(\varphi+\chi); \quad \eta=u_a\sin(\varphi+\chi). \quad (23)$$

The deviation of the ray passing through the deformed lens in the plane of focus ($z=z_F$) is then[11]

$$\Delta u=\frac{f}{a}\int_{-\infty}^{\zeta_F}Ru_ad\zeta. \quad (24)$$

Here, to obtain a first approximation, the coordinates in Eq. (23) are to be substituted for u, ξ, and η in the expression $R(\zeta)$ from Eq. (21). It is desirable to refer this displacement to the plane in which the corresponding ray for the symmetrical lens traverses focus. The coordinates are then given by

$$\Delta\xi_\varphi+i\Delta\eta_\varphi=\Delta ue^{-i(\varphi-\chi_a)}; \quad \chi_a=-\int_{-\infty}^{\zeta_F}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi'd\zeta. \quad (25)$$

χ_a is the total rotation of the image. $\Delta\xi_\varphi$ now represents the deviation parallel to the meridional plane, $\Delta\eta_\varphi$ that at right angles thereto. Substituting R from Eq. (21) in Eq. (24),

$$\Delta\xi_\varphi+i\Delta\eta_\varphi=-\frac{2if}{a}e^{i\chi_a}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\int_{-\infty}^{\zeta_F}\left[\Delta\Psi'\left(u_a'u_a-i\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi'u_a^2\right)\right.$$
$$\left.-e^{-i(\varphi+\chi)}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}\cos(\varphi+\chi)+\Delta\frac{\partial^2\psi}{\partial\eta^2}\sin(\varphi+\chi)\right)u_a^2\right]d\zeta. \quad (26)$$

$$\Delta\xi_\varphi=\frac{f}{a}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\left\{2\sin\chi_a\int_{-\infty}^{\zeta_F}\frac{\Delta\Psi'}{\Psi_{\max}'}u_au_a'd\zeta\right.$$
$$-2\cos\chi_a\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\int_{-\infty}^{\zeta_F}\frac{\Psi'\Delta\Psi'}{\Psi_{\max}'^2}u_a^2d\zeta+\sin\chi_a\int_{-\infty}^{\zeta_F}\frac{u_a^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)d\zeta$$
$$-\cos2\varphi\int_{-\infty}^{\zeta_F}\frac{u_a^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\sin(2\chi-\chi_a)d\zeta$$
$$\left.-\sin2\varphi\int_{-\infty}^{\zeta_F}\frac{u_a^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\cos(2\chi-\chi_a)d\zeta\right\}. \quad (27)$$

[11] See e.g., E. P. Adams, *Smithsonian Mathematical Formulae* (Smithsonian Institution, Washington, D. C., 1922), Formula 8.410.

$$\Delta\eta_\varphi=\frac{f}{a}\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\Big\{-2\cos\chi_a\int_{-\infty}^{\zeta_F}\frac{\Delta\Psi'}{\Psi_{\max}'}u_a u_a'd\zeta$$

$$-2\sin\chi_a\left(\frac{e}{8mc^2\Phi}\right)^{\frac{1}{2}}\Psi_{\max}'\int_{-\infty}^{\zeta_F}\frac{\Psi'\Delta\Psi'}{\Psi_{\max}'^2}u_a^2d\zeta-\cos\chi_a\int_{-\infty}^{\zeta_F}\frac{u_a^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)d\zeta$$

$$-\cos2\varphi\int_{-\infty}^{\zeta_F}\frac{u_a^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}-\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\cos(2\chi-\chi_a)d\zeta$$

$$+\sin2\varphi\int_{-\infty}^{\zeta_F}\frac{u_a^2}{\Psi_{\max}'}\left(\Delta\frac{\partial^2\psi}{\partial\xi^2}+\Delta\frac{\partial^2\psi}{\partial\eta^2}\right)\sin(2\chi-\chi_a)d\zeta. \quad (28)$$

The distance between the focal plane and the plane in which the ray approaches closest to the axis is given by

$$\Delta z_F=-a\Delta\xi_\varphi/u_a'=f\Delta\xi_\varphi; \quad (29)$$

$ar_0\Delta\eta_\varphi$, r_0 signifying the distance of the ray from the axis in image space, on the other hand, gives the deviation from the axis at the point of closest approach. Equations (27) and (28) were evaluated by numerically solving Eq. (22) for u_a and performing the indicated quadratures. The results are shown in Table I and Fig. 18.

APPENDIX II

Compensation of Objective Asymmetries

Equation (7), representing the axial potential for any pole-piece assembly with two meridional planes of symmetry passing through the ξ and η axes, may be written

$$\psi=\Psi_0+\left(\frac{\partial^2\psi}{\partial\xi^2}\right)\xi^2+\left(\frac{\partial^2\psi}{\partial\eta^2}\right)\eta^2+\cdots$$

$$=\Psi_0+\frac{1}{2}\left[\frac{\partial^2\psi}{\partial\xi^2}+\frac{\partial^2\psi}{\partial\eta^2}\right](\xi^2+\eta^2)+\frac{1}{2}\left[\frac{\partial^2\psi}{\partial\xi^2}-\frac{\partial^2\psi}{\partial\eta^2}\right](\xi^2-\eta^2)+\cdots. \quad (30)$$

The last term in this expression is responsible for the astigmatism of imaging pencils refracted by the system.

Assume now that a second asymmetry is introduced, e.g., by a pair of iron screws placed symmetrically with respect to the axis. Let the principal axes of this "compensating" system (ξ_1, η_1) form an angle Θ with the ξ, η axes:

$$\xi_1=\xi\cos\Theta-\eta\sin\Theta,\quad \eta_1=\xi\sin\Theta+\eta\cos\Theta,\quad \xi_1^2-\eta_1^2=\xi^2\cos2\Theta-2\xi\eta\sin2\Theta-\eta^2\cos2\Theta. \quad (31)$$

The total asymmetry term resulting from the superposition of the two fields is now

$$\Delta\Psi=A[\xi^2(1-k\cos2\Theta)-2k\xi\eta\sin2\Theta-\eta^2(1-k\cos2\Theta)], \quad (32)$$

where

$$A=\frac{1}{2}\left(\frac{\partial^2\psi}{\partial\xi^2}-\frac{\partial^2\psi}{\partial\eta^2}\right);\quad k=\left(\frac{\partial^2\psi_1}{\partial\xi_1^2}-\frac{\partial^2\psi_1}{\partial\eta_1^2}\right)\Big/\left(\frac{\partial^2\psi}{\partial\xi^2}-\frac{\partial^2\psi}{\partial\eta^2}\right).$$

The subscript 1 refers to the compensating system. If the factor k, normally a function of ζ, is a constant, perfect compensation is possible. This occurs, of course, for $\Theta=0$ and $k=1$ and corresponds to a compensating field whose asymmetry terms are, except for sign, identical throughout with those of the original field.

If the factor k is not a constant, the adjustment of Θ and the magnitude of k (for a short magnetic or for an electrostatic lens) will still bring about correction, though now the imagining pencils can be made stigmatic only in the first order. The image formation in such systems is analogous to the formation of an image focused sharply in two perpendicular directions by two cylindrical lenses with their axes at right angles to each other. For a long magnetic lens compensation will generally be incomplete, since the adjustment of the orientation and the magnitude of an arbitrary compensating field will not suffice to balance both the isotropic and the anisotropic component of the astigmatism resulting from the original field. Below the case of the short lens, for which a single, effective, value of k can be defined, will be treated.

Since, in practice, the orientation of the asymmetry in the original pole pieces is initially either unknown or known only approximately, the principal axes of the compensating field first introduced will, in general, form an angle Θ different from zero with the principal axes of the original field. It is hence of interest to know how the asymmetry in the image will change with the value of k for different values of Θ.

For any value of k and Θ the principal axes (ξ_2, η_2) of the resultant field (determining the directions of maximum focal deviation or most pronounced defocusing fringes in the image) will be rotated relative to the principal axes ξ, η of the original field by an angle ϑ:

$$\xi = \xi_2 \cos\vartheta - \eta_2 \sin\vartheta, \quad \eta = \xi_2 \sin\vartheta + \eta_2 \cos\vartheta. \tag{33}$$

Substituting this relation in Eq. (32) and adjusting ϑ so that the coefficient of $\xi_2\eta_2$ vanishes leads to

$$\Delta\Psi = A(1 - 2k\cos 2\Theta + k^2)^{\frac{1}{2}}(\xi_2^2 - \eta_2^2), \tag{34}$$

$$\tan 2\vartheta = -k \sin 2\Theta/(1 - k\cos 2\Theta). \tag{35}$$

Here the coefficient in Eq. (34) measures the asymmetry of the resultant field (or the value of the eccentricity factor ϵ^2), ϑ the rotation of the direction of maximum lens action (or of the orientation of the most pronounced fringes) as k (or the compensating field) is increased from zero to the given value. These quantities are plotted, for a series of values of Θ, the angular deviation between the principal axes of the two superposed fields, in the polar diagram shown in Fig. 33. Each curve corresponds to a range in k from 0 to 2; arrowheads indicate the direction of increasing k.

Correction takes place only if Θ is less than 45° in absolute value. If this condition is fulfilled, increasing the compensating field from zero results in an initial reduction in the degree of asymmetry and a subsequent increase. The value of ϑ passes simultaneously from 0 to $\mp(\pi/2 - |\Theta|)$ if the sign of Θ is $\pm$. Only for $\Theta = 0$ is full correction attained; for $|\Theta| = 5°$ the asymmetry can be reduced to less than a fifth, for $|\Theta| = 10°$, to a third, and for $|\Theta| = 15°$, to a half.

An Investigation of a Method for the Analysis of Smokes According to Particle Size[1]

WILLIAM N. LIPSCOMB,[2] T. R. RUBIN,[3] AND J. H. STURDIVANT
California Institute of Technology, Pasadena, California[4]

(Received August 2, 1946)

An electrical precipitation method for the analysis of smokes according to particle size has been investigated. A thin stream of smoke particles moving under laminar conditions in a wind tunnel of rectangular cross section is electrically charged in a small region. The charged particles are then precipitated by the uniform field existing between the charged upper plate and the grounded lower plate of the wind tunnel as a long track along the lower plate. The size-frequency distributions of the particles precipitated at various points on the lower plate were determined from measurements on enlargements of electron-microscope photographs of samples of the smoke precipitated at various points. The agreement between theory and experiment is fair, but the range of sizes of the particles precipitated at a given point is rather large.

INTRODUCTION

THE use of a precipitation method for the estimation of sizes of smoke particles was first described by Rohmann.[5] His apparatus consisted of a wind tunnel of rectangular cross section with an electric field between the negatively charged upper plate and the grounded lower plate. The smoke was introduced into the wind tunnel through a small tube and was charged in a small region by a corona discharge from a point. The charged particles were precipitated by the uniform field as a track about 5 cm in length. The size of the particles was measured with an optical microscope and compared with the particle radius calculated from the theory of precipitation of charged spherical particles in a uniform field. Rohmann did not report the size-frequency distributions of the particles precipitated at various points, but only the average size in the densest region of the precipitate. No statement was made regarding the regularity of shape of the metallic oxide particles in the smokes which he used for his investigation. Rohmann mentions two difficulties encountered in his experiments. The first was the difficulty of obtaining laminar gas flow in the wind tunnel; even in the absence of a charging corona he was forced to work with gas velocities of nearly 40 cm/sec. in order to obtain a stable thread of smoke. The second difficulty was the disturbance of the thread of smoke by the electrical wind which accompanies corona discharge; he was therefore constrained to use comparatively small charging currents.

The present investigation consists of a more extensive study of the precipitation method for the analysis of smokes according to particle size. A modified form of the apparatus described by Rohmann was used. The rectangular wind tunnel was redesigned in order to give essentially laminar flow at gas velocities of approximately 10 cm/sec. Since the point corona was found to be unsatisfactory, other methods of obtaining charging currents were investigated. The electron microscope was used to record the sizes of particles precipitated at various points in the apparatus. Had the separation of the smoke particles proved more satisfactory a much more rapid method was to be used (based on calibration with the electron microscope results) for the determination of the size-frequency distribution function; this proposed method consisted of measurement of the current carried by the charged smoke particles to small metal collectors mounted in the plane of the lower plate and insulated from it. Because the results were not completely satisfactory, however, only the studies of various charging devices, representative size-frequency distribution curves, and com-

[1] This work was done in whole under the Contract No. OEMsr-103, Supplement 1, between the California Institute of Technology and the Office of Scientific Research and Development, which assumes no responsibility for the accuracy of the statements contained herein.
[2] Present address: School of Chemistry, University of Minnesota, Minneapolis 14, Minnesota.
[3] Present address: Department of Chemistry, Ohio State University, Columbus, Ohio.
[4] Contribution No. 1054.
[5] H. Rohmann, Zeits. f. Physik **17**, 253 (1923).

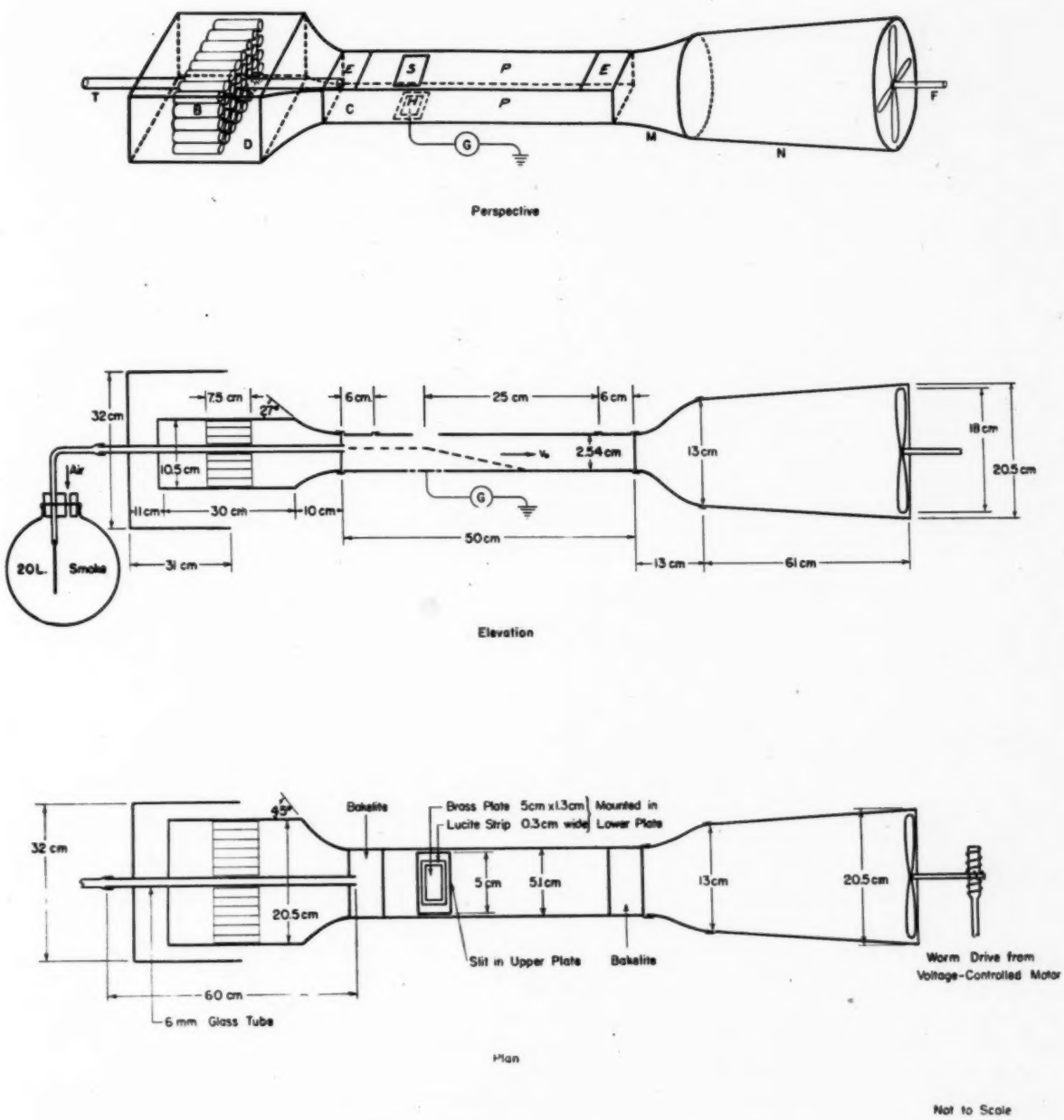


FIG. 1. The precipitation apparatus.

parisons of experimental results with theoretical precipitation curves are presented here.

APPARATUS

A diagram of the precipitation apparatus is shown in Fig. 1. The smoke, which was contained in a large flask, was forced by means of compressed air first through a short length of small-bore glass tubing (to reduce turbulence) and then into the entrance tube *T*. This tube extended through the entrance duct *D* to the rectangular precipitation chamber *C*. In order to minimize turbulence the entrance duct was provided with a honeycomb baffle *B* made from short lengths of Cellophane soda straws. Further reduction of turbulence was obtained by placing a box in front of the entrance duct in the position indicated in the plan and elevation of Fig. 1. The

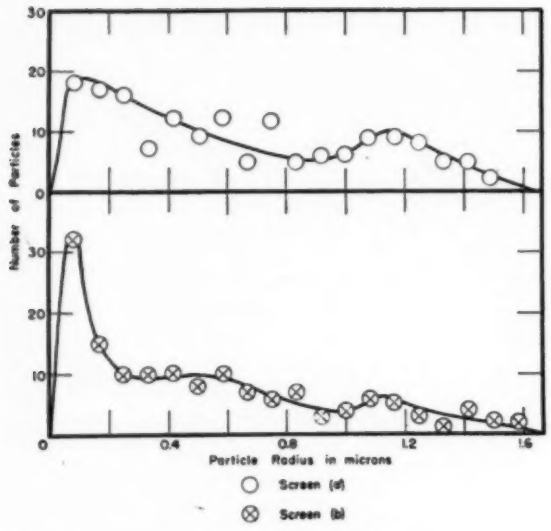


FIG. 2a. Size-frequency distributions of particles in a rosin smoke charged by means of a heated filament.

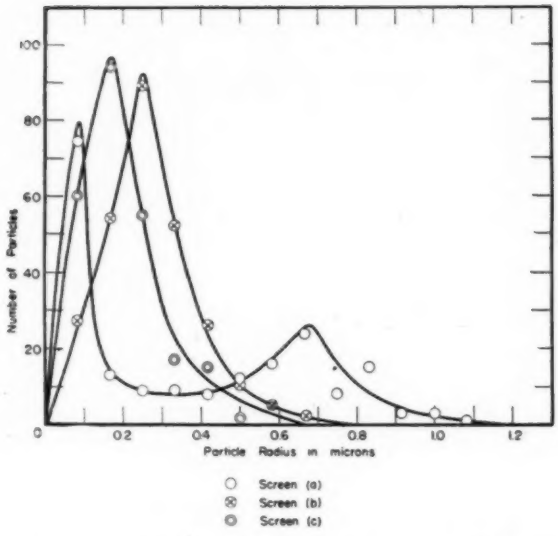


FIG. 3. Size-frequency distributions of particles in a rosin smoke charged by means of a cotton cloth at high potential.

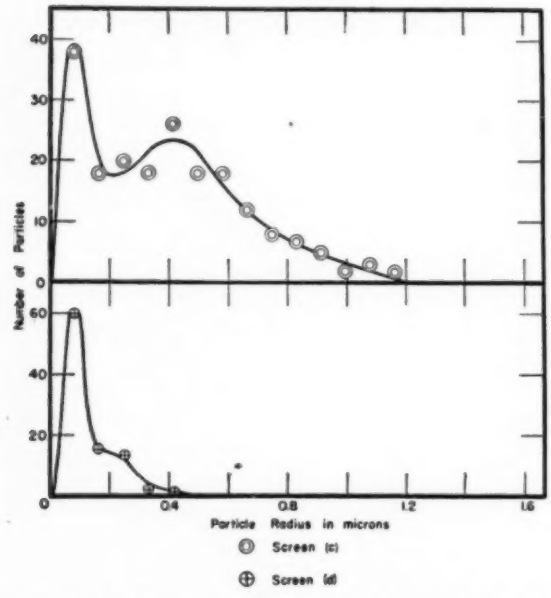


FIG. 2b. Size-frequency distributions of particles in a rosin smoke charged by means of a heated filament.

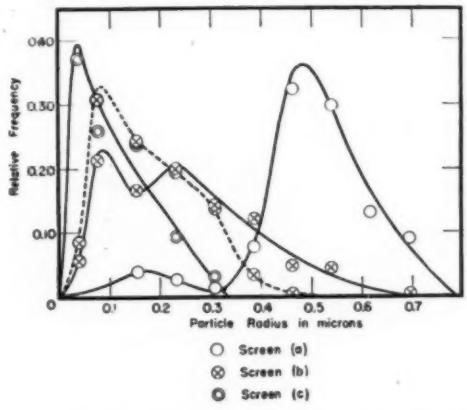


FIG. 4. Size-frequency distributions of particles in a rosin smoke charged by means of a plate coated with rayon flock fibers.

precipitation chamber consisted of two parallel brass plates P separated by lucite strips. The upper plate, which was charged to about -4 kv, was insulated from the entrance and exit ducts by bakelite strips E. The various charging devices which were tested were placed above or in the plane of an aperture S in the upper plate, and were surrounded by an air-tight insulator to prevent air from entering around the charging region. The lower plate was provided with a metallic collector H which was mounted in the plane of this plate and insulated from it. This collector was made movable along the direction of wind flow in order to measure, by means of a galvanometer, the iron currents emitted by the charging device as a function of the position of the collector. The air-stream outlet consisted of two sections which ended finally in a large circular cross-section. A four-bladed fan, run from a variable speed motor, was placed just inside the outlet section N; this fan produced air velocities at the center of the precipitation

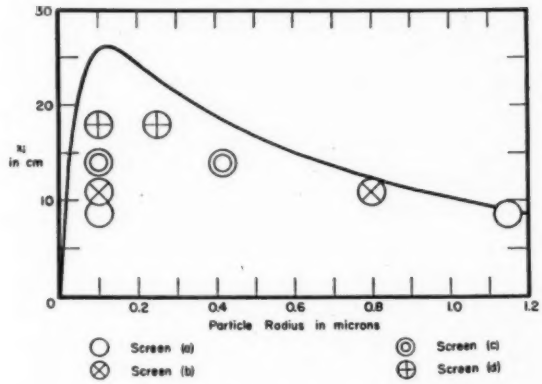


FIG. 5. Average particle radius as a function of point of precipitation; charging device was a heated filament.

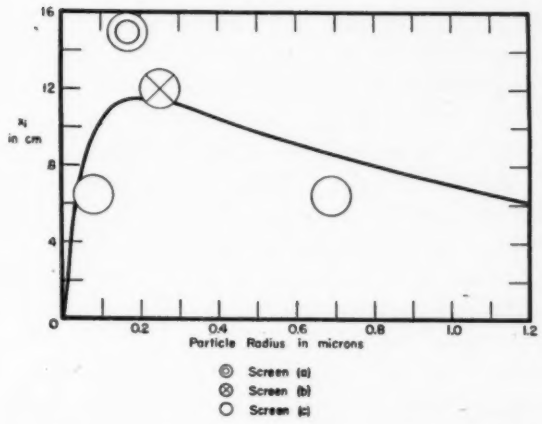


FIG. 6. Average particle radius as a function of point of precipitation; charging device was a cotton cloth.

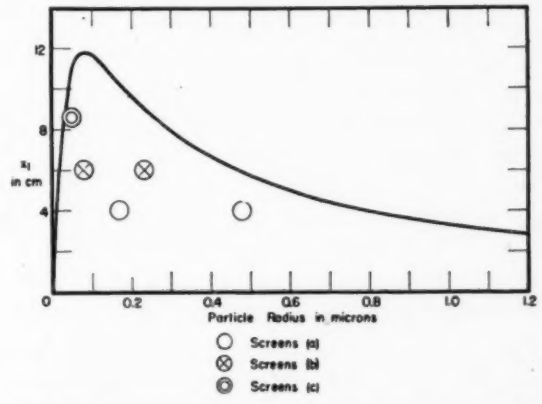


FIG. 7. Average particle radius as a function of point of precipitation; charging device was made from rayon flock fibers.

chamber of 8 to 25 cm/sec. The velocities which were used in the runs described below were 12 to 15 cm/sec.

A uniform downward current of ions of about 10^{-8} ampere/cm² must be obtained in order to charge the smoke particles. Of the large number of devices which were tested for this purpose, the three most promising were investigated in detail by studies of the size-frequency distribution of the smoke particles precipitated at various points in the apparatus. The first (Run 1) consisted of a small heated Nichrome coil which was supported by a short quartz rod and jacketed by a Pyrex compressed air cooler. Immediately below this coiled filament was placed a wire screen (15–50 meshes/in.²) which was allowed to remain at "floating" potential. The filament was heated by means of a filament transformer and was maintained at a potential of about −5 kv. The second charging device (Run 2) consisted of a small area (1 cm×2.5 cm) of cotton bird's-eye toweling which was placed in contact with a metal plate at −4 to −10 kv. The third charging device (Run 3) consisted of a small aluminum plate maintained at −14 kv and covered sparsely with rayon flock fibers (about 250/cm²) attached

TABLE I. Summary of experimental results.

Screen	Particles measured	x_i in cm	Position of maxima in microns
Run 1; Charging device: heated filament			
(a)	176	8.7	0.1 1.15
(b)	146	11.0	0.1 0.5, 1.1 (Average, 0.8)
(c)	195	14.0	0.1 0.42
(d)	99	18.0	0.1 0.25
Run 2; Charging device: cotton cloth			
Screen	Particles measured	x_i in cm	Position of maxima and half width at half maximum, in microns
(a)	196	6.5	0.08±0.05, 0.69±0.13
(b)	265	12.0	0.25±0.08
(c)	248	15.0	0.17±0.10
Run 3; Charging device: rayon flock fibers			
Screen	Particles measured	x_i in cm	Position of maxima in microns
(a)	77	4.0	0.16 0.48
(b)	229	6.0	0.08 0.23
	452	6.0	0.09 (0.2 to 0.3)
(c)	97	8.6	0.05

normal to it.[6] This third charging device gave the most nearly reproducible and uniform charging conditions of all of those which were tested.

For studies with the electron microscope, collodion-covered screens were placed on the lower plate of the precipitation apparatus at various distances from the charging region. These screens were 6-mm circles cut from 350-mesh brass screen, and were coated with a collodion film about 0.1 micron in thickness.

EXPERIMENTAL PROCEDURE

The wind velocity v_0 was measured in the following manner: a thin stream of smoke was allowed to flow down the center of the precipitation chamber. A disturbance was given to the stream by a slight movement of the inlet tube, and the time required for this disturbance to move a known distance was measured.

The ion current in the charging region was measured at various positions by means of the movable collector plate; it was found to be fairly uniform and limited to a small region. The extent of the charging region, extrapolated to the height of the smoke stream above the lower plate, divided by the velocity of the stream was taken as the time of charging, t. The average ion current per unit area in the charging region was taken as the ion current density, i.

The rosin smokes were prepared by heating about 70 mg of rosin in a small copper pan; the rosin condensed into a smoke consisting of approximately spherical particles of radius $0.05 < a < 3$ microns. The smoke was collected in the large flask and was forced into the apparatus by a slow stream of compressed air. When the charging and deflecting potentials were turned on, the smoke stream was observed to curve down slightly and fan out into a curtain of smoke dropping to the lower plate. The smoke thus deposited a narrow track of rosin on the lower plate where, for the purpose of electron microscopy, the small collodion-covered collectors were placed.

[6] Flock is a preparation of textile fibers cut to uniform, short lengths. In order to prepare this charging device the required density of flock was placed on a grounded conducting plate. The aluminum plate, wet with Glyptal, was suspended above the grounded plate and charged to about −15 kv in order to deposit the fibers in an erect position on the charged plate. The potential was left on until the Glyptal hardened.

Photographs were taken of the smoke deposits on the collodion-covered screens with the use of the electron microscope designed at the California Institute of Technology by Professor W. V. Houston and Dr. J. H. Bradner, and the RCA electron microscope at the Department of Physics of the University of California at Berkeley. Enlargements of the electron microscope photographs were made to a total linear magnification of about 6000. For spherical particles the diameter was measured; for elliptical and slightly irregular particles the average of the largest and smallest diameters was taken. Nearly all of the rosin smoke particles were approximately spherical except for the larger particles obtained in Run 1, in which the heated filament was used as a charging device. These highly irregular particles were divided into spherical or approximately elliptical sections and measured as independent parts. Although this idealization is clearly not a good approximation, a convenient method of comparing particle sizes was obtained by this method. There is no satisfactory treatment of the law of fall for particles of irregular shape. From the measurements of particle size, the size-frequency distribution curves were prepared by plotting the number of particles of a given size as ordinate against the radius of the particle in microns as abscissa, and then drawing a smooth curve through the points.

RESULTS

The size-frequency distribution curves obtained in the three runs are shown in Figs. 2a, 2b, 3, and 4. The experimental results are summarized in Table I. On Screen (b) of Run 1 two maxima appeared for $a > 0.3$ microns whereas the theory required only the one shown by most of the other experimental results. The average of these two peaks was taken. It is to be noted that the numbers of particles of given size determining parts of this and some of the other distribution curves are too small to yield satisfactory statistical distributions. As a result the method of interpolation by which some parts of the curves should be obtained is not unambiguous. The photographed area of Screen (b) of Run 3 was divided into two parts, and the particles on these two parts were counted separately in order to determine whether the result-

ing size-frequency distribution curves were in satisfactory agreement. The results, shown in Fig. 4 and Table I are in good agreement, but it is quite clear that a large number of particles must be counted before satisfactory distributions are obtained.

The experimental results are compared with the theoretical curves, calculated as described in the Appendix, in Figs. 5, 6, and 7. The numerical expressions from which the curves were calculated are, for Run 1,

$$x_i=\frac{1660a_\mu}{(1+0.082/a_\mu)Q} \text{ where } Q=123a_\mu^2+23a_\mu;$$

for Run 2,

$$x_i=\frac{1840a_\mu}{(1+0.082/a_\mu)Q} \text{ where } Q=157a_\mu^2+82a_\mu;$$

and, for Run 3,

$$x_i=\frac{1840a_\mu}{(1+0.082/a_\mu)Q} \text{ where } Q=476a_\mu^2+39a_\mu,$$

where a_μ is the particle radius in microns. The agreement between the curves and the experimental points is not too good. However, the general character of the theoretical curves is reproduced by the experiments, for example, the precipitation of both large and small particles at a given point as required by the theory appears to be verified.

In order to test whether the original smoke sample itself contained an abnormally large number of small particles, the whole thread of smoke was swept repeatedly across a collodion-covered screen by variation of the corona potential, until a satisfactory density of particles was obtained. Electron microscope photographs of this sample did not show a predominant number of particles of radius less than 0.2 micron; the size-frequency distribution curve for the unseparated smoke showed a single maximum at about 1 micron.

Considerable experimental difficulties were encountered in the use of the various charging devices. The filament heated the charging region considerably, and even though approximate thermal equilibrium had been reached before the run was started some slight thermal unsteadiness in the precipitation conditions persisted. The cotton cloth charging device produced results which were quite unreproducible; however, no variations in the charging conditions were usually observed during a single run. The conducting plate covered with the rayon flock fibers was probably the most satisfactory charging device which was tested. These fibers undoubtedly retained their relative orientation for successive runs better than did the cloth fibers. The appearance of the track of rosin precipitated on the lower plate was very similar for successive runs with the flock fiber charging device, although a series of five successive runs under as nearly the same conditions as possible gave results which agreed rather poorly.

The reasons for this poor agreement probably include (a) variations in the initial charges on the particles, (b) turbulence in the wind flow in the tunnel, (c) irregularities in the particle shape, (d) fluctuations in the charging current, and (e) statistical variations in the charging. The effects of variation of the initial charges of the particles were estimated by measurement of the maximum deflection of the smoke stream in a uniform field in the absence of the precipitating current. The maximum deflection observed corresponded to a maximum initial charge of $\pm 8e$ on a particle of radius 0.5 micron, or $\pm 34e$ on a particle of radius 2 microns. Turbulence is extremely difficult to eliminate completely, and its effect was apparent on careful visual observation under the best conditions which we obtained; it may have been caused, in part, by electrical wind effects. Fluctuations in the charging current of 1 to 2 percent were detected with a probe which was connected to an electrometer circuit. The effect of irregularities in the shapes of the particles and of statistical variations in charging could not be estimated.

ACKNOWLEDGMENT

Our thanks are due Professor Linus Pauling, who suggested this study, for his interest, advice, and encouragement. We are also indebted to Mr. George Standart for assistance with part of the experimental work, and to Dr. R. V. Langmuir and Mr. R. Rau of the California Institute of Technology, Mr. A. E. Smith of the Shell Development Company, and Mr. A. W. Einarsson of the University of California at Berkeley

for their assistance in helping us to obtain electron microscope photographs.

APPENDIX: THEORETICAL SECTION

A particle which has a radius a and charge Qe and is swept along a horizontal rectangular channel (see Fig. 8) between two charged plates has a downward velocity v' because of the gravitational and electrostatic forces, $mg+E_0Qe$;

$$v'=\frac{mg+E_0Qe}{6\pi\eta a}\left(1+\frac{A\lambda}{a}\right), \tag{1}$$

where η is the viscosity of the gas; A is the Cunningham constant[7]; and λ is the mean free path of the gas molecules.

The problem of the gas velocity at any point in a tube of rectangular cross section has been solved by Cornish.[8] In a tube with cross-sectional width c and height d in which the velocity on the axis ($y=z=0$) is v_0, the velocity of gas flow along any line parallel to the axis is

$$v=v_0\left[\frac{\frac{1}{4}-\frac{y^2}{d^2}-\frac{8}{\pi^3}\left(\frac{\cosh\frac{\pi z}{d}}{\cosh\frac{\pi c}{2d}}\cos\frac{\pi y}{d}-\frac{1}{3^3}\frac{\cosh\frac{3\pi z}{d}}{\cosh\frac{3\pi c}{2d}}\cos\frac{3\pi y}{d}+\cdots\right)}{\frac{1}{4}-\frac{8}{\pi^3}\left(\frac{1}{\cosh\frac{\pi c}{2d}}-\frac{1}{3^3\cosh\frac{3\pi c}{2d}}+\cdots\right)}\right] \tag{2}$$

The horizontal distance, x_i, from the point at which the particles are charged to the point at which they strike the lower plate may then be shown, by integration of the equation $dx=(v/v')dy$, to be

$$x_i=\frac{6\pi\eta a v_0}{(mg+E_0Qe)\left(1+\frac{A\lambda}{a}\right)}\frac{2l^2}{d}\left(1-\frac{2l}{3d}\right)\times\left\{1+\frac{32}{\pi^3\cosh\frac{\pi c}{2d}}\left[1+\frac{32}{\pi^3\cosh\frac{\pi c}{2d}}\right]\times\left[1-\frac{d^2\left(1-\cos\frac{\pi l}{d}\right)\cosh\frac{\pi z}{d}}{2\pi l^2\left(1-\frac{2l}{3d}\right)}\right]+\cdots\right\}, \tag{3}$$

where l is the height above the lower plate at which the stream of particles is charged. For the conditions under which the experiments were carried out $E_0Qe\gg mg$ and $z=0$, and an adequate approximation to Eq. (3) is

$$x_i=\frac{6\pi\eta a v_0}{E_0Qe\left(1+\frac{A\lambda}{a}\right)}\frac{2l^2}{d}\left(1-\frac{2l}{3d}\right). \tag{4}$$

This equation would be suitable for calculation if we knew the number of charges, Q, as a function of the particle radius and known or measurable quantities.

A theory of the charging of a spherical particle of radius a and dielectric constant K passing through a downward current of ions has been derived by Rohmann[5] under the assumptions that (1) diffusion of charges to the particle is negligible, (2) electrical image forces between an ion approaching a particle and the particle are negligible, and (3) all of the incident ions give up their charge to the particle. His expression for the charge Q_fe acquired in time t by an initially uncharged particle may be written in the form

$$Q_fe=\left(1+\frac{2(K-1)}{K+2}\right)E_1a^2\frac{\pi it}{E_1+\pi it}, \tag{5}$$

where i is the ion current per unit area and E_1 is the field intensity in the charged region.

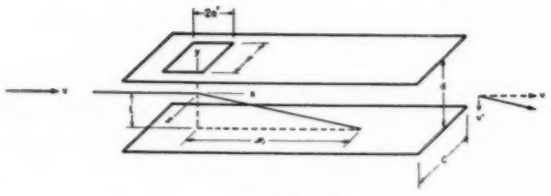


FIG. 8.

No satisfactory theory has been developed for the initial stage of the charging of a spherical particle of low dielectric constant by diffusion. The final stage of diffusion charging was investigated theoretically and experimentally by Arendt and Kallmann.[9] Their work, and also the experimental work of Deutsch[10] and Schweitzer[11] showed that the saturation charge taken up by oil drops of radius from 0.1 micron to 10 microns in a space charge of the order of 10^5 charges/cm^3 is proportional to the particle radius. Ladenburg[12] pointed out that in the process of precipitation of smoke particles in a charging field there are two distinct charging mechanisms, one, important for the larger particles, giving a charge proportional to a^2 as described in the preceding paragraph, and another, important for the smaller particles, giving a charge proportional to a. This second mechanism is called diffusion charging although it includes, in addition to ordinary diffusion of ions to the particle, the effect of image forces on the charge approaching the dielectric surface of the particle and the effect of the net charge on the diffusion of ions to the particle. We have assumed that the number of charges diffusing to a particle of radius a in a time t and in an ion atmosphere of volume density f is

$$Q_d=Cfta. \tag{6}$$

The value of the proportionality constant was determined to be $C=0.07$ by extrapolation to zero time of the experimental curves for the charging of oil drops as a function of time, as given by Arendt and Kallmann.[9] Unfortunately, Arendt and Kallmann did not obtain data for times less than about 4 min., and consequently the constant C cannot *a priori* be regarded as determined even to the correct order of magnitude for the time intervals of interest ($t\sim0.2$ sec.). We shall, nevertheless, use the resulting expression for calculation of Q_d because our results described above indicate that the resulting term in the

[7] E. Cunningham, Proc. Roy. Soc. **A83**, 357 (1910).
[8] R. J. Cornish, Proc. Roy. Soc. **A120**, 691 (1928).
[9] P. Arendt and H. Kallmann, Zeits. f. Physik **35**, 421 (1926).
[10] W. Deutsch, Zeits. f. tech. Physik **7**, 623 (1926).
[11] H. Schweitzer, Ann. d. Physik [5] **4**, 33 (1930).
[12] R. Ladenburg, Ann. der Physik [5] **4**, 863 (1930).

charging law is approximately correct, and is almost certainly of the correct order of magnitude. Since the ion density is given by the relation $f=i/(E_1ue)$, where u is the mobility of the ions, Eq. (6) therefore becomes

$$Q_de=0.07iat/E_1u. \tag{7}$$

In the absence of a complete mathematical treatment of the charging process, we shall write the expression for the total charge at a time t on a spherical particle of radius a in the form

$$Qe=Q_fe+Q_de, \tag{8}$$

where Q_fe and Q_de are given by Eqs. (5) and (7), respectively. In this and the preceding equations the effect of the comparatively small initial charge on the particle has been omitted. Although the two charging processes are probably not independent, and the expressions for the two processes therefore not strictly additive, it may be pointed out that because of the different dependence of the two processes on the particle radius, Eq. (7) represents the important term for the smaller particles, whereas Eq. (6) represents the important term for the larger particles.

The value of Qe from Eq. (8) may be substituted into Eq. (4) in order to calculate the distance along the direction of gas flow from the center of the charging region to the point of precipitation of a particle of radius a. In this calculation we have made the additional assumption that the charging, which occurs over the distance $v_0t=2a'$ (Fig. 8), may be approximated by complete charging at a point at the center of the charging region.

In some of the experiments the field in the charging region was considerably higher than that between the deflecting plates of the rectangular precipitation chamber. The average field in the charging region in this case was calculated from an approximate solution of the two-dimensional potential distribution by expansion in rectangular harmonics. For the case (Run 3, above) in which this treatment was applied, the field thus calculated was about 7 percent smaller than that obtained from the relation $E_1=-V_1/d$, where V_1 is the potential of the upper plate of the rectangular tube. The effect of space charge in decreasing the field intensity in the charging region was small enough to be neglected.

Dislocation Theory as Applied by N.A.C.A. to the Creep of Metals

A. S. Nowick and E. S. Machlin
Aircraft Engine Research Laboratory, Cleveland, Ohio

(Received September 9, 1946)

An equation for the steady-state rate of creep of pure annealed polycrystalline metals is derived through dislocation theory and the theory of rate processes. The rate of generation of dislocations is shown to be the rate-determining process. A specific mechanism for the generation of a dislocation is presented, from which the heat and entropy of activation of the process are expressed in terms of physical constants of the material. In addition the lowering under stress of the potential energy barrier, which yields the stress-dependent term, is given in terms of constants of the material and a "back-stress" term. The creep equation thus obtained is found to be in good agreement with data in the literature.

INTRODUCTION

CREEP of metals is generally subdivided into three stages, as shown in Fig. 1. The initial stage, in which the slope of the elongation-*versus*-time curve (the rate of creep) is rapidly decreasing, is commonly designated the primary or transient stage; secondary or steady-state creep refers to the straight-line portion of the creep curve. After a sufficient length of time, the rate of creep increases in the region designated the tertiary stage.

The criterion generally used to evaluate the creep property of a given material at a particular stress and temperature is the steady-state creep rate. One of the first attempts to analyze steady-state creep by other than empirical methods was by Kanter.[1] Later, Kauzmann[2] applied Eyring's theory of liquid flow, which is a special application of the theory of rate processes, to the steady-state flow of metals. The problem of creep has also been attacked by the use of the theory of dislocations. A survey of results obtained thus far by this approach is reported by Seitz and Read.[3] Although these three treatments predict equations for the steady-state creep rate that show the correct dependence on stress and temperature, quantitative predictions cannot be

[1] J. J. Kanter, Trans. A.I.M.E. **131**, 385 (1938).
[2] W. Kauzmann, Trans. A.I.M.E. **143**, 57 (1941).
[3] F. Seitz and T. A. Read, J. App. Phys. **12**, 100, 170, 470, 538 (1941).

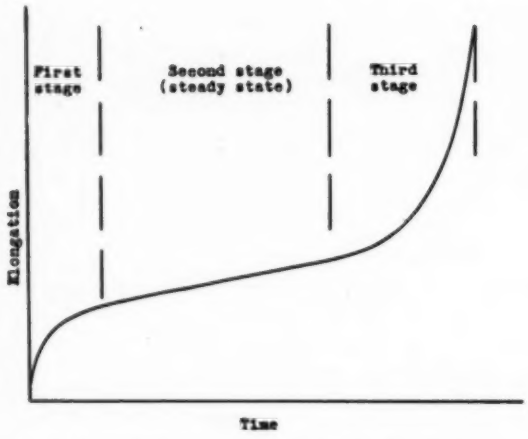


FIG. 1. A typical creep curve showing the three stages.

made from them as to how physical and structural constants of materials affect creep rate. Furthermore, difficulties were encountered in attempting to explain the large negative entropy of activation that experimental data showed.

This report derives an equation for the steady-state creep rate in terms of a specific dislocation mechanism which involves considering the generation of dislocations to be a rate process. The features of the theory of rate processes and of dislocation theory required for the development of the creep equation are first briefly reviewed; the derivation then follows.

The material contained herein was derived at the Cleveland laboratory of the N.A.C.A. and was originally presented as N.A.C.A. Technical Note No. 1039.

THEORY OF RATE PROCESSES

The theory of rate processes,[4] as developed by Eyring and others, considers a reaction or any rate process to be the result of the crossing of a potential-energy barrier by molecules whose energies have exceeded a certain minimum value. The reaction consists of the formation of an "activated complex" capable of crossing the barrier followed by the passage of this complex over the barrier. The most important assumption is that the initial reactants and activated complexes are always in equilibrium. Application of thermodynamic and statistical mechanical considerations shows that the number r of activated complexes passing over the barrier per second (or rate of reaction), per unit concentration of reactants is given by

$$r=\frac{kT}{h}\exp(-\Delta F_a/kT) \qquad (1)$$

in terms of Boltzmann's constant k, Planck's constant h, the absolute temperature T, and the free energy of activation per molecule, ΔF_a.

The term kT/h can be regarded as the effective frequency at which activated complexes cross over the barrier; the exponential factor represents the probability of formation of an activated complex. In the calculation of this probability term, the contribution due to the translational degree of freedom along the "reaction coordinate" (the most favorable reaction path on the potential-energy surface) has been disregarded because it is included in the factor kT/h. The free energy of activation ΔF_a is interpreted as an ordinary free-energy term and can be expressed by

$$\Delta F_a=\Delta H_a-T\Delta S_a, \qquad (2)$$

where the heat of activation per molecule ΔH_a is the height of the potential-energy barrier, and ΔS_a is the entropy of activation per molecule.

DISLOCATION THEORY IN RELATION TO SLIP

Of several mechanisms suggested, the one that has been most successful in explaining the slip process is the theory of dislocations. This theory proposes that local deviations from a perfect lattice, called dislocations, exist in single crystals and that movement of dislocations through the stressed crystal produces slip. A dislocation consists in a stable arrangement of atoms such that, in a region of a few atomic distances, $n+1$ atoms in the slip direction face an adjacent parallel row of n atoms across the slip plane. The point at which the atoms are one-half an atomic spacing "out of step" is called the center of the dislocation. The extension of the dislocation in the slip plane and normal to the slip direction is called the length of the dislocation. Two types of dislocations can exist: a dislocation is called positive if in the neighborhood of the center of the dislocation the compression is above the slip plane, and negative if the compression occurs

[4] S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes* (McGraw-Hill Book Company, Inc., New York and London, 1941), Chaps. I, IV, and IX.

below the slip plane. Positive and negative dislocations move in opposite directions under a given shear stress, but they produce the same net result in passing through the crystal, a relative translation of one atomic distance between the parts of the lattice above and below the slip plane.

Inasmuch as the passing of a dislocation out of the crystal lattice results in its loss, new dislocations must be generated in order for continuous plastic deformation to be possible. Because of the high energy required, the process generally takes place at regions of high stress concentration resulting from imperfections in the single crystals. The existence of such imperfections as a result of accidents of growth during the formation of single crystals is generally accepted, but the nature of the flaws that exist is still a matter of controversy. For example, the theory that single crystals are made up of a mosaic block structure is discussed by Seitz and Read. X-ray evidence indicates that the average spacing of the imperfections is of the order of 1 micron. The dynamic dislocation picture obtained is as follows: dislocations are generated under shear stress at the regions of high stress concentration, which will be called sources of dislocations. Upon generation, the dislocations are only a few atoms long because of the extremely low probability of generating a full-length dislocation (one whose length is of the order of the spacing between imperfections). As the dislocations move through the crystal under a shear stress, they increase in length and move in the slip plane and slip direction until they become stuck at an imperfection or in the vicinity of other dislocations.

Further details on the properties of dislocations as used to explain the qualitative aspects of plastic deformation are given by Seitz and Read.[3]

DERIVATION OF THE CREEP EQUATION

In the analysis of the creep process, it will be assumed that creep as well as slip takes place by the motion of dislocations. If L is the spacing between imperfections (in terms of the mosaic structure theory, the length of a mosaic block), we choose a cubic block of side L and make two assumptions: (a) no "annihilation" (union of pairs of dislocations of opposite sign) takes place so that each dislocation passes completely through the block; (b) the block chosen is representative of the entire crystal in that it will shear at the same rate as any other similar block. This assumption follows from the requirement of a steady-state condition.

For the chosen block, the number of dislocations generated per second within the block must equal the number leaving the block per second, or

$$R_g N_s L^3 = N_d v L, \tag{3}$$

where R_g is the rate of generation for a single generating source; N_s is the number of sources per unit volume; N_d is the number of dislocations per unit area that intersect a plane normal to the slip plane and containing the slip direction; and v is the average velocity of the moving dislocations. Inasmuch as the displacement produced when a dislocation passes out of the block is equal to the lattice spacing d_1 between atoms in the slip direction, the shear rate u' is given by

$$u' = \frac{1}{L}(N_d v L) d_1 = N_d v d_1. \tag{4}$$

The occurrence of slip bands at distances apart of the order of 1 micron indicates an average of about one source of dislocations for each block so that

$$u' = (d_1/L) R_g. \tag{5}$$

Up to this point, only creep of single crystals has been considered. In polycrystalline metals, it would appear that two types of creep are possible: creep within individual grains, and creep resulting from intergranular motion. As noted by Seitz and Read, present data on the effect of grain size on creep leave doubt as to the possibility of the occurrence of intergranular creep. For example, it has been shown[5] in work on copper of various grain sizes that there are no detectable differences in creep strength other than those resulting from oxidation. (Unfortunately this work was carried out only at a single temperature.) Creep in polycrystalline metals is therefore assumed to be predominately the result of deformation occurring within the individual grains by the motion of dislocations. If the

[5] E. R. Parker and C. F. Riisness, Metals Tech. A.I.M.E. 11 (1944), Tech. Pub. No. 1690, p. 1.

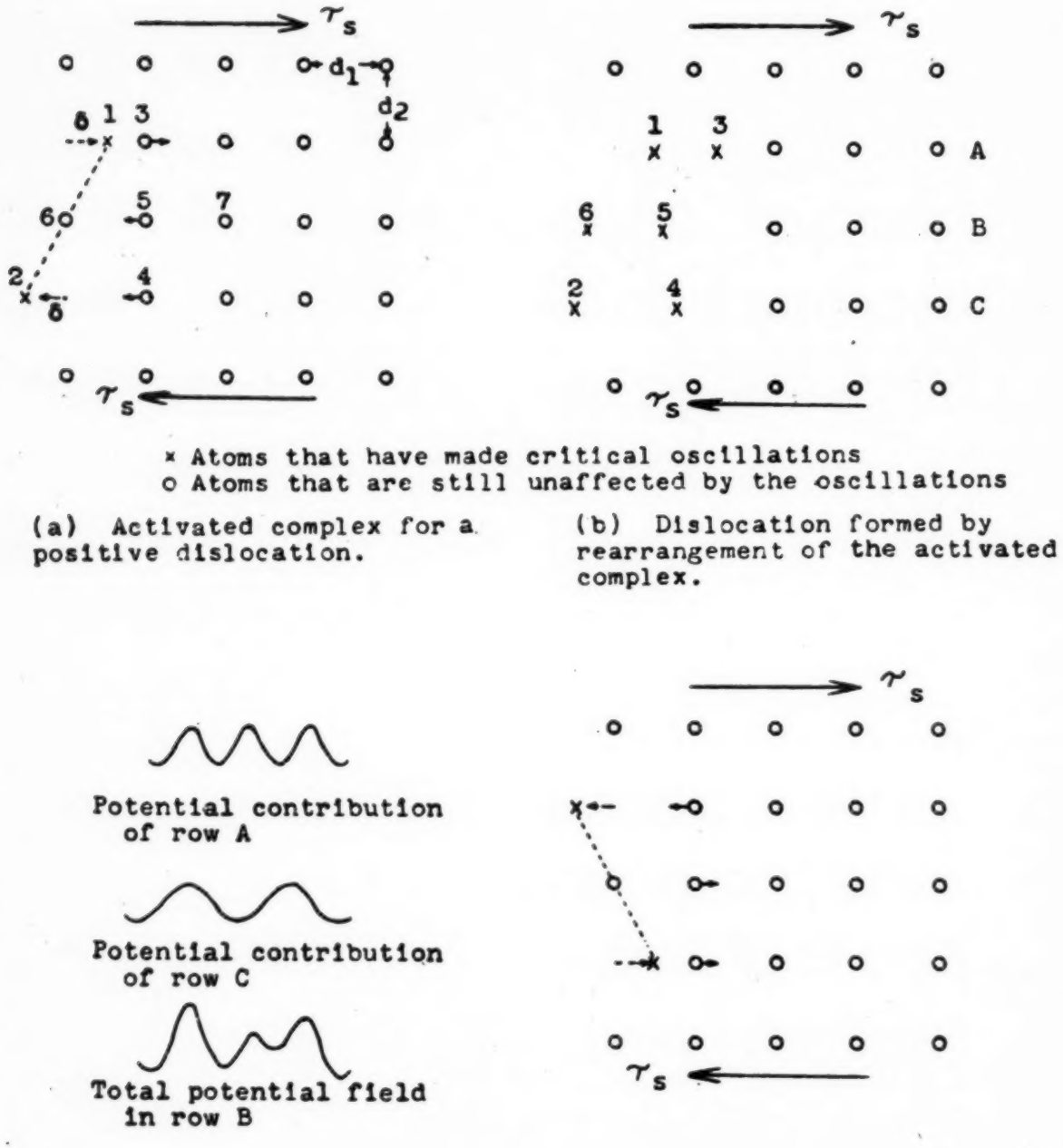


FIG. 2. Generation of a dislocation at a positive source.

orientation of the grains is such that slip occurs in the plane of maximum shear stress, then, following Kauzmann, the strain rate u' can be corrected to the tensile creep rate u by an additional factor of about $\frac{2}{3}$. Such a factor would be of no consequence here because d_1/L is known only in order of magnitude; accurate values cannot be given until more is known about the nature of the crystal imperfections. The tensile creep rate u of a polycrystalline material can

therefore be written approximately as

$$u=(d_1/L)R_g. \tag{5'}$$

In calculating the rate of generation R_g, the generated dislocation will be taken to be 1 atom long. A simple model for the generation of a positive dislocation at a right-hand source[6] is presented in Figs. 2(a) and (b). If atoms 1 and 2 are simultaneously displaced in the directions shown in Fig. 2(a) by large fractions of the lattice spacing, and atoms 3, 4, and 5 are moving in the slip direction as shown, the generation of a dislocation becomes possible. Before a dislocation will form, however, a rearrangement of the neighboring atoms must take place in order to achieve a stable configuration. Figure 2(b) shows the rearranged condition. The new positions of atoms 5 and 6 after rearrangement were found by using the approximate method of Taylor[7] in his discussion of the motion of a dislocation under stress. The contributions to the potential field in row B by the neighboring rows A and C are taken to be sinusoidal with periods equal to the spacing of atoms 1 and 3 in row A and atoms 2 and 4 in row C. Figure 2(c) shows that addition of the two sinusoidal potential functions gives a resultant potential function having four minima. Atoms 5, 6, and 7 can be expected to fall into the deepest of these minima if atom 5 had begun by oscillating to the left as in Fig. 2(a).

The displacement δ of atoms 1 and 2 will be some fraction f of the atomic separation d_1 in the slip direction. A value of f of somewhat less than $\frac{1}{2}$ makes it possible to form a stable configuration with a minimum activation energy.

The mechanism just presented is an example of a rate process where the "reactant" is the perfect lattice at the region of high stress concentration. The "activated complex" for a positive dislocation, which for brevity will be called a positive activated complex, can be taken as the configuration shown in Fig. 2(a) and the "product" as the dislocation after atomic rearrangement has taken place (Fig. 2(b)). If the fundamental assumption of rate-process theory that equilibrium exists between reactant and activated complex is made, the considerations involved in the theory of rate processes can be applied to the generation of dislocations. If t is the average time for the activated complex to pass into final form, the rate of generation of positive dislocations R_g^+ can be expressed as

$$R_g^+=(1/t)P^+, \tag{6}$$

where P^+ is the probability of formation of a positive activated complex at a right-hand source.

It has been shown[8] that if the displacement involved in the transition from the activated complex to the final state is on the order of 10^{-8} centimeter, the factor $1/t$ is approximately equal to kT/h. In the formation of a dislocation, this transition corresponds to passing from the configuration of Fig. 2(a) to that of Fig. 2(b), which clearly involves a displacement of this order of magnitude. Hence, R_g^+ is given approximately by

$$R_g^+=(kT/h)P^+. \tag{7}$$

Similarly, for the rate of generation of negative dislocations,

$$R_g^-=(kT/h)P^-, \tag{7'}$$

where P^- is the probability of formation of a negative activated complex at a right-hand source.

In order to simplify the calculation of P^+ and P^-, the generation of a dislocation will be treated as a "local shear" which takes place by means of thermal oscillations. The energy involved in this process will be calculated by the theory of elasticity. Furthermore, Hooke's law will be assumed valid for large strains. The use of these assumptions is necessary because the forces between individual atoms have not been evaluated in terms of physical constants, but it should be remembered that the approximations may be serious enough to lead to large errors in the results. The thermal atomic oscillations in Fig. 2(a) will be regarded as resulting in a local shearing of atoms 1, 6, and 2 producing a strain of δ/d_2. Inasmuch as the elastic energy per unit volume for a shear stress τ is $\tau^2/2G$ where G is the modulus of rigidity, and the volume of the

[6] A source of dislocations will be called a right-hand source if it generates dislocations to the right, and a left-hand source if it generates to the left. The sources are taken to be at "internal surfaces," for example, the edge of a surface of misfit between mosaic blocks.

[7] G. I. Taylor, Proc. Roy. Soc. (London) **A145**, 362 (1934).

[8] Glasstone, Laidler, and Eyring, reference 4, p. 189.

region of shear under consideration is twice the atomic volume V, the elastic energy involved is

$$W = V\tau^2/G. \tag{8}$$

Now, let τ_0 be the local shear stress attained by atoms 1, 6, and 2 in Fig. 2(a). If the displacement δ is some fraction f of d_1, application of Hooke's law to the displacement yields

$$\tau_0 = Gxf, \tag{9}$$

where x is defined as the ratio d_1/d_2. If the source of generation is under a stress τ_s, as shown in Fig. 2, the energy W_t to be attained thermally is, by (8)

$$W_t = (V/G)(\tau_0 - \tau_s)^2. \tag{10}$$

Inasmuch as τ_s may be expected to be only a small fraction of τ_0 for rigid materials, the term in τ_s^2 may be neglected in the binomial expansion of $(\tau_0 - \tau_s)^2$, so that, incorporating (9), we obtain

$$W_t = VGx^2f^2 - 2Vxf\tau_s. \tag{11}$$

In order to evaluate τ_s, work-hardening must be considered. A crystal in the steady-state range of creep is generally work-hardened because deformation has taken place. From the nature of a dislocation it is clear that a stress field exists about it. By making use of this stress field, Taylor[7] has explained work-hardening on the basis of a "lattice" of stuck dislocations. Taylor's explanation is a static one and is applied only to determining the form of the stress-strain curve. In terms of a dynamic approach, the existence of a stress field about a dislocation leads to the concept suggested by Kochendörfer[9] that the lattice of stuck dislocations creates an inner stress field whose direction at the point of generation is opposite to the externally applied shear stress τ_e in the slip plane. The value of the inner stress field at the point of generation will be called the back stress τ_b. The dynamic concept of hardening therefore consists in the formation of a lattice of stuck dislocations, which results in a back stress at the point of generation. This back stress lowers the effective shear stress at these points, and as a result the rate of generation, hence the creep rate, decreases. These considerations lead to the result that

$$\tau_s = q(\tau_e - \tau_b), \tag{12}$$

where q is the stress concentration at the generating source. Inasmuch as τ_b is always less than τ_e, the directions of τ_e and τ_s are the same.

In addition to the energy of local shear, the configuration of Fig. 2(a) involves a limitation of the movement of atoms 1, 2, 3, 4, and 5 to a particular crystallographic direction within a very small solid angle. If the probability of oscillation in this direction is p, the total probability of the configuration of Fig. 2(a) is

$$P^+ = p^5 \exp(-W_t/kT) = p^5 \exp[-(VGx^2f^2 - 2Vxf\tau_s)/kT]. \tag{13}$$

Similarly the probability of forming a negative activated complex (the configuration of Fig. 2(d)) at a positive source is

$$P^- = p^5 \exp[-(VGx^2f^2 + 2Vxf\tau_s)/kT], \tag{13'}$$

because τ_s is opposite to τ_0 in this case. The applied stress τ_e therefore makes the generation of a positive dislocation more probable than the generation of a negative dislocation at a right-hand source.

If a negative dislocation is generated in the region under discussion, it would move to the left and out of the crystal without producing deformation if it were the only dislocation present. The last positive dislocation generated from this source, however, became stuck after moving through part of the crystal lattice; that is, all forces upon it were balanced. The generation of a negative dislocation at the point being considered will upset this equilibrium and attract the previous positive dislocation. As the two dislocations approach each other, the stress that each exerts on the other increases,[10] and a union takes place resulting in the annihilation of the two dislocations and restoration of a perfect lattice. The net rate of generation R_g of positive dislocations that will eventually produce plastic deformation is $R_g^+ - R_g^-$ or from Eqs. (7), (7'), (12), (13), and (13')

$$R_g = \frac{2kT}{h} \exp[-(VGx^2f^2 + CT)/kT] \times \sinh[2qVxf(\tau_e - \tau_b)/kT], \tag{14}$$

where

$$C = -k \ln p^5. \tag{15}$$

[9] A. Kochendörfer, Zeits. f. Physik **108**, 244 (1938).

[10] From the equation for the stress field about a dislocation; see Taylor, reference 7.

The equation for the net rate of generation of negative dislocations at a negative source is clearly identical with (14).

According to this derivation, it is now clear that the generation of a dislocation 2 atoms in length involves the simultaneous oscillation of twice as many atoms with sufficiently great energies. Thus, the probability of such a generation taking place can be neglected as compared with the probability of generating a 1-atom dislocation.

Comparison of (14) with Eqs. (1) and (2) now shows that the heat of activation (the height of the barrier for zero applied stress) and the entropy of activation are given, respectively, by

$$\Delta H_a = VGx^2f^2, \tag{16}$$

$$\Delta S_a = -C. \tag{17}$$

C is always positive because p is less than 1, so ΔS_a is negative.

When the expression (14) for R_a is combined with the creep equation (5′), the complete equation for the steady-state creep rate u becomes

$$u = \frac{2d_1}{L}\frac{kT}{h}\exp[-(VGx^2f^2+CT)/kT] \times \sinh[qVxf(\sigma-2\tau_b)/kT], \tag{18}$$

where the applied tensile stress σ has been substituted for $2\tau_e$ assuming that slip in the individual grains occurs in the plane of maximum shear stress. For high values of σ, the hyperbolic sine function becomes an exponential, or in logarithmic form

$$\ln u = \ln\left(\frac{d_1kT}{Lh}\right) - \left(\frac{VGx^2f^2+CT}{kT}\right) + \frac{qVxf}{kT}(\sigma-2\tau_b). \tag{19}$$

COMPARISON OF THEORETICAL AND EXPERIMENTAL CREEP EQUATIONS

The evaluation of the back stress τ_b as a function of T and σ is difficult and has not yet been accomplished. We will therefore compare the theoretical results with empirical equations. It has been found that a hyperbolic sine function represents the dependence of creep rate on applied tensile stress σ for many polycrystalline materials[11] in agreement with Eq. (18). It has also been shown[12] that creep rates of a large number of materials can be represented in the form

$$\ln\frac{u}{T} = -\frac{a}{T} - b + \gamma\sigma, \tag{20}$$

where a and b are constants for a given material and γ is a function of temperature but not of stress. If the notation is changed, (20) can be rewritten

$$\ln u = \ln\left(\frac{d_1kT}{Lh}\right) - \frac{A+BT}{kT} + \frac{\beta\sigma}{kT}, \tag{21}$$

where

$$A = ka,$$

$$B = kb + k\ln\frac{d_1k}{Lh},$$

$$\beta = \gamma kT.$$

The simultaneous validity of (19) and (21) requires that τ_b be a linear function of stress. In addition, the condition $\tau_b = 0$ for $\sigma = 0$ which should apply at least to annealed metals, yields

$$\tau_b = F(T)\sigma, \tag{22}$$

where $F(T)$ depends only on temperature and material parameters. Equation (21) shows that three factors determine the creep rate, the constants A and B and the temperature function β. Comparison of (21) with (19) shows that these factors are expressed in terms of physical constants of the material as

$$A = VG_0f^2x^2, \tag{23}$$

$$B = C - Vf^2x^2G_0\alpha, \tag{24}$$

$$\beta = qVxf(1-2F(T)), \tag{25}$$

where the modulus of rigidity G has been taken in the approximate form

$$G = G_0(1-\alpha T), \tag{26}$$

where G_0 is the modulus of rigidity at absolute zero.

It seems clear that the structure sensitivity of the creep property, which is particularly important in alloys, will be mainly due to the back

[11] P. G. McVetty, Trans. A.S.M.E. **65**, 761 (1943).
[12] S. Dushman, L. W. Dunbar, and H. Huthsteiner, J. App. Phys. **15**, 108 (1944).

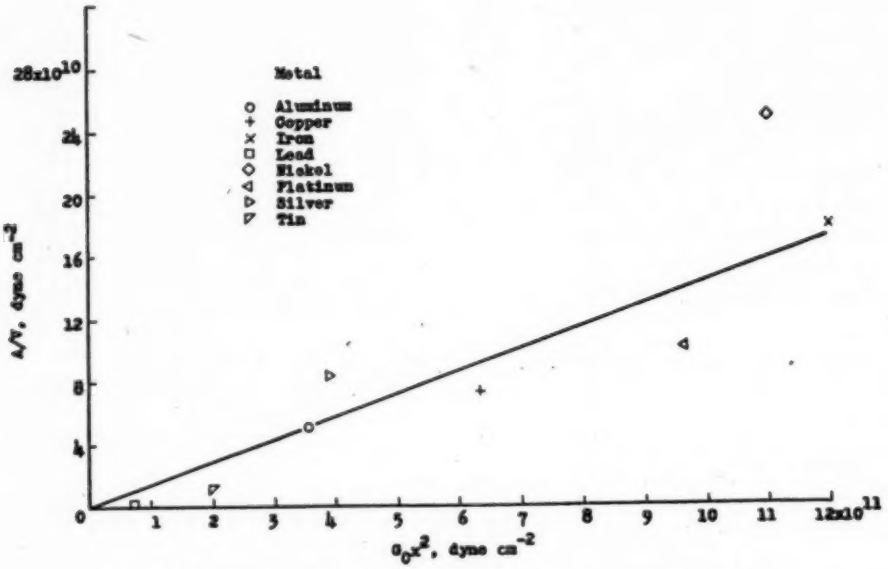


FIG. 3. Comparison of experimental data for factor A with theory.

stress term and will therefore appear in the factor β.

COMPUTATIONS AND PRECISION

Creep data were obtained from the literature[13] for a number of polycrystalline "pure" metals that had been annealed. The reproducibility of the creep data from the various sources is not known. The data were plotted up similarly to Dushman and the factors A, B, and β obtained. The factor d_1/L was taken as 10^{-4} for all materials.

Generally, the errors in A, B, and β were estimated to be about $\pm 5\times 10^{-13}$ erg, $\pm 8\times 10^{-16}$ erg/°K, and ± 30 percent, respectively.

Room temperature values of G were used as approximations to G_0.

RESULTS AND DISCUSSION

The Factor A

In Fig. 3, A/V is plotted against G_0x^2. The line drawn is the best straight line through the origin and generally represents the data within the estimated error of A. From Eq. (23) and the above discussion the slope of the line is f^2 and should be somewhat less than 0.25. Because the activation energy predicted on the basis of the validity of Hooke's law will be greater than the actual value, it can be expected that the theoretical prediction will be too high. The experimental slope from the straight line of Fig. 3 is 0.14. This agreement is good considering the assumptions made and variety of sources of data used. It gives further verification that the generated dislocation is only one atomic distance in length.

The Factor B

The order of magnitude of the term $Vf^2x^2G_0\alpha$ of Eq. (24) which is 3×10^{-16} erg/°K, is within the experimental error of B, $\pm 8\times 10^{-16}$ erg/°K; this term can therefore be neglected in consider-

TABLE I. Results for the parameter B.
(The values are given in units of 10^{-16} erg/°K.)

Metal	B
Aluminum	40
Copper	45
Iron	39
Lead	54
Nickel	33
Platinum	42
Silver	36
Tin	47

[13] Aluminum, platinum, and silver: Dushman *et al.*, reference 12; copper: E. A. Davis, J. App. Mech. **10**, A101 (1943); H. L. Burghoff and A. I. Blank, Trans. A.I.M.E. **161**, 420 (1946); nickel: A. Michel and J. Cournot, Congrès International pour l'Essai des Matérieux (Amsterdam) **1**, 397 (1927); iron: A.S.T.M.—A.S.M.E., *Compilation of Available High-Temperature Creep Characteristics of Metals and Alloys* (1938), p. 22; zinc: W. M. Peirce and E. A. Anderson, Trans. A.I.M.E. **83**, 560 (1929); tin and lead: L. C. Tyte, Proc. Phys. Soc. **50**, 153 (1938); *ibid.* **51**, 203 (1939).

ation of the B values. Therefore B becomes equal to C, or from (17) to the negative of the entropy of activation ΔS_a. The results obtained from the data are listed in Table I.

A correlation exists between the variation from the best straight line of data presented in Fig. 3 and the variation in B values. The points above the line are all related to metals that have B values less than 40 and those below the line to B values greater than 40. From the method of plotting the data to obtain A and B values, a positive error in A means a negative error in B. This correlation therefore indicates that the correct values of B fall about their mean value more closely than the values in the table. In addition, errors in the substitution of 10^{-4} for d_1/L and kT/h for $1/t$ will enter into the B factor. The entropy of activation ΔS_a therefore seems to be very closely the same for all materials considered. This is in agreement with the theory since the factor p should be about the same for all materials.

The large negative value of the entropy of activation ΔS_a has previously been pointed out by Kauzmann and by Dushman, but satisfactory explanations for the magnitude of this factor could not be given. In terms of the present treatment, the large negative ΔS_a is accounted for by the restriction of direction of 5 atoms participating in the generation of an activated complex. A value of about 3×10^{-3} is obtained for p when $C=40\times10^{-16}$ erg/°K. The possibility that more than 5 atoms must be restricted in direction certainly exists, so that p may be larger than the above value.

The Factor β

For a large number of materials the temperature dependence of β could be approximated by the function

$$\beta=C_1 \exp C_2 T,$$

where C_1 and C_2 are positive constants. This function was used to extrapolate to room temperature values.

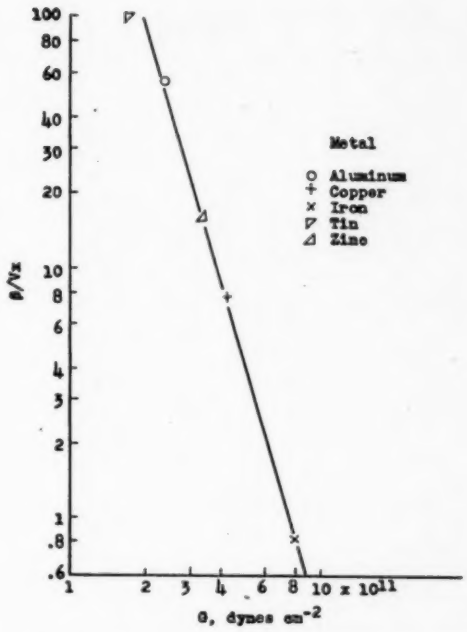


FIG. 4. Room temperature dependence of β/Vx on the modulus of rigidity G.

The term $F(T)$ in (25) is related to the back stress τ_b which results from the internal stress field about a dislocation. This stress field is dependent on G. Therefore, β at any given temperature was expected to be a function of G. Figure 4 shows that, for the materials shown, the β values extrapolated to room temperature vary as G^{-n} where n is approximately 3.

CONCLUSIONS

The analysis presented in this report shows that dislocations are generated by the formation of an activated complex configuration in a small region. Limitation of the direction of motion of a group of atoms results in a large negative entropy of activation for the process. When generated, the dislocations are probably just one atom long.

Application of the Sliding Thermocouple Method to the Determination of Temperatures at the Interface of a Moving Bullet and a Gun Barrel*

E. L. Armi,† J. L. Johnson, R. C. Machler, and N. E. Polster
Leeds and Northrup Company, Philadelphia, Pennsylvania

(Received May 31, 1946)

Description is given of a set-up for the determination of temperatures at the frictional interface between a moving bullet and the bore in a caliber 0.50 gun barrel. Cathode-ray oscilloscope e.m.f.-time records, which were obtained from a traveling thermocouple junction formed by the bullet-bore interface, are interpreted (by use of a formula due to Jaeger) in terms of instantaneous temperatures. A calorimetrically measured value of the total frictional heat input caused by a bullet being pushed through a barrel section by means of a falling weight, is found to be in good agreement with the value of the heat input calculated from bore surface temperatures which were determined by the sliding thermocouple method.

INTRODUCTION

DURING the course of an investigation on values of local heat input associated with the firing of individual rounds of ammunition in caliber 0.50 machine gun barrels, it became desirable to obtain an idea of the magnitude of that part of the heat which is caused by friction between bullet and bore, and also an idea of the bore surface temperatures which are reached during this frictional process, before the onset of heat input from the hot powder gases.

A falling weight apparatus with a usable free fall of 65-foot length was modified to drive bullets with fair velocity through a short length of barrel held in a suitable mounting. Making use of the fact that the frictional interface between bullet and bore is also an e.m.f.-generating interface, because of a difference in the thermoelectric potential of the two materials involved (gun steel, gilding metal), effective temperatures at the interface were determined from a measurement of the e.m.f. from this somewhat complex "sliding band thermocouple." Having found the approximate temperature-time curve of a point on the bore, it was easy to calculate the local frictional heat input into the barrel wall. On the other hand, this heat input was also measured directly by a calorimetric method, and thus a check was obtained on the results of the sliding thermocouple method.

Sliding thermocouple measurements were also carried out on caliber 0.50 barrels in actual firings, but the check against calorimetric heat input measurements had to be omitted in this case, because of the impossibility of separating the frictional heat input from that of the powder gases.

FALLING WEIGHT APPARATUS

Figure 1 gives a view of the falling weight apparatus, with the carriage in raised position. Two taut wire rope guides reached up to a height of 65 feet above a solid concrete foundation. The carriage, sliding on four bearings, was pulled up by a rope and was dropped with the aid of an electromagnetic release mechanism. The weight of the carriage, including the steel bar at its bottom, amounted to 40 lb; occasionally, additional bars were clamped onto the carriage in order to increase the kinetic energy of the system. At the end of its fall the carriage hit a piston which in turn drove the bullet through a short length of barrel, held in a heavy steel base. Excess kinetic energy was taken up by lead stops of suitable size.

Electrical connection to the bullet was made by means of a shielded flexible cable soldered to the tip of the bullet; the grounded shield of the cable was connected to the barrel section. E.m.f.'s generated at the frictional interface were amplified in a Ballantine Decade Amplifier (Model 220) before being made visible on the screen of a DuMont cathode-ray oscillograph

* The work reported in this article was undertaken as part of contract OEMsr-536 between the Leeds and Northrup Company and the Office of Scientific Research and Development, under the supervision of Division One of the National Defense Research Committee.

† Now Director of Technical Department, Mica Insulator Company, Schenectady, New York.

(Type 175A). The latter was operated with single sweeps which were obtained from a DuMont low-frequency linear time base Generator (Type 215) and which were triggered by the closing of a contact shortly before the carriage hit the piston driving the bullet. Figure 1 shows the double spike on the carriage, which closed the sweep initiating circuit of the time base generator by connecting the two mercury pools in the reservoirs visible below the spikes.

Details of the construction of the barrel section holder can be seen in Fig. 2. (The thermocouple shown on the outside of the barrel section was used for the calorimetric measurements discussed under the heading "Theory of Calorimetric Measurements in Barrel Sections.") In most tests the barrel sections had special 0.500″/0.520″ rifling (bore diameter 0.500″ across lands, 0.520″ across grooves) instead of the regular 0.500″/0.510″ rifling of caliber 0.50 guns, in order to avoid contact of the bullets with the bottom of the grooves of the rifling; the generation of thermal e.m.f. was thereby restricted to the area of the top and (part of) the sides of the rifling bands ("lands"), and the interpretation of the records obtained became somewhat simpler. All barrel sections had at their entrance a forcing cone (origin of rifling) of standard taper.

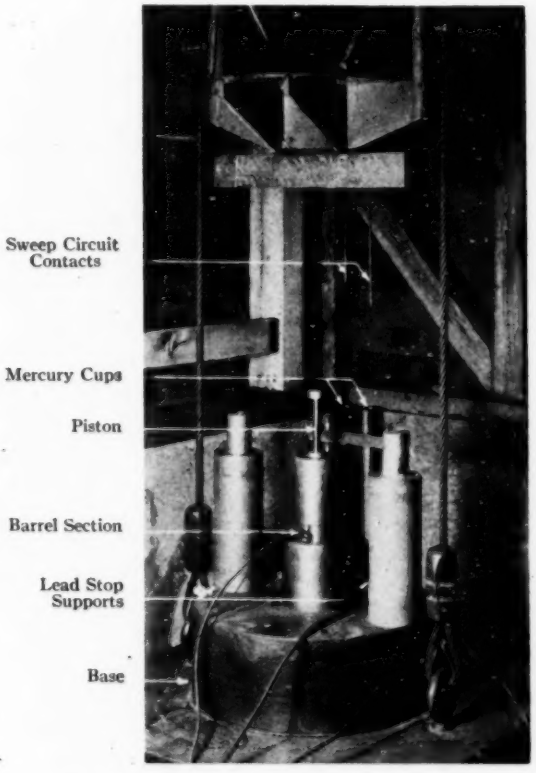


FIG. 1. Falling weight apparatus (carriage in raised position).

INTERPRETATION OF CRO TRACES OF THE INTERFACE THERMAL E.M.F.

Interpretation of the experimental data was greatly helped by a paper of J. C. Jaeger, on "Moving Sources of Heat and the Temperatures at Sliding Contacts,"[1] in which temperatures in a semi-infinite conductor are worked out for the motion of a band source of heat over the conductor, and this analysis is applied to the case where a slider moves with constant friction over a metallic surface. Jaeger shows that, for high

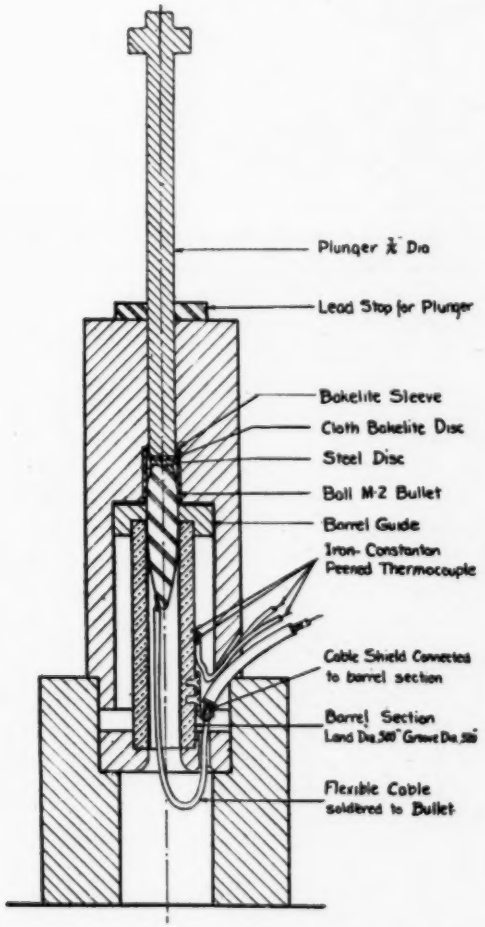


FIG. 2. Details of mounting of barrel section.

[1] J. C. Jaeger, Proc. Roy. Soc., N. S. W. **76**, 203–224 (1943).

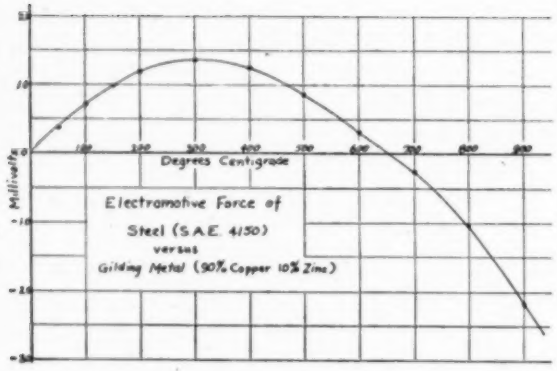


FIG. 3. E.m.f. calibration curve of gun steel *vs.* gilding metal.

sliding speeds and after a steady state is reached, the instantaneous temperature distribution over the surface of the metal in contact with the slider can be approximated by[2]

$$\theta/\theta_{\max} = (L/L_0)^{\frac{1}{2}}, \qquad (1)$$

where θ is the temperature rise on the surface of the metal, L the distance back from the front edge of the "band source," and L_0 the total width of the "band source." It seems permissible, for a first approach to the solution of the problem, to assume that the conditions of the theory are satisfied to an acceptable degree in the case of a bullet sliding over the (spirally rifled) bore surface, whereby the frictional contact area corresponds to the heat-generating band of the mathematical analysis.

Matters become complicated by the behavior of the e.m.f.-temperature curve of the particular combination of metals in the sliding thermocouple (gun steel *vs.* gilding metal). Figure 3 shows a calibration curve taken on a sample thermocouple, the two wires of which were cut from a bullet jacket and a gun barrel, respectively. This curve, which has its maximum at about 300°C, is parabolic in its general shape; all e.m.f. readings are, therefore, essentially ambiguous since their correlated temperatures can lie either on the ascending or the descending branch of the calibration curve. It should also be kept in mind that the e.m.f. recorded is essentially an arithmetical average of the e.m.f.'s generated over the whole frictional interface area at the instant concerned; but, because of the non-linearity of the e.m.f.-temperature curve, the temperature corresponding to the average of the e.m.f.'s is not at all identical with the average of the temperatures which cause the generation of these e.m.f.'s.

Figure 4, based on Eq. (1) and Fig. 3, gives curves of the e.m.f.'s which are generated locally along the frictional interface. Of course, these e.m.f.'s are never measured as such, since only the average over all locally generated e.m.f.'s is recorded on the CRO screen. Figure 5, above the designation "full contact," shows this average for several values of $\theta_{\max}$. However, it should be remembered that the "local" e.m.f.'s are actually averages too, namely over-all points which make contact with the bore on a particular circumference of the bullet. (It is therefore, in general, quite possible to find points which have temperatures in excess of the effective "maximum temperature.") Since the temperature in the frictional process is unlikely to exceed the melting point of the material of the bullet jacket (gilding metal), the curves for a fictitious $\theta_{\max} = 1200°$ have to stop at the e.m.f. corresponding to the temperature of that melting point (about 1050°C).

As long as the bullet is fully in contact with the bore surface, averaging of the e.m.f.'s takes place over the whole of the curves of Fig. 4. However, when the bullet is entering or leaving the barrel section, the average is taken over only that part of the curve which corresponds to actual contact with the bore. Figure 5 gives e.m.f.

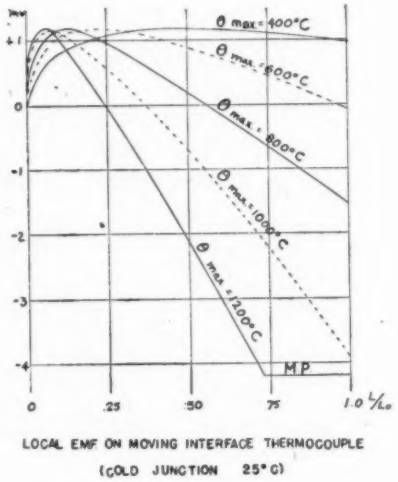


FIG. 4.

[2] J. C. Jaeger, reference 1, Eq. (14).

curves which were obtained by carrying out the averaging as indicated, for the bullet entering, traversing, and leaving the barrel section, with X/L_0 denoting the ratio between instantaneous length of contact and length of contact for a fully entered bullet. The relative length of the center part of the curve, designated "full contact," varies, of course, with the length of the barrel section used.

COMPARISON BETWEEN THEORETICAL AND OBSERVED FORMS OF OSCILLOGRAM RECORDS

Figure 6 shows three frictional interface e.m.f. traces, together with theoretical curves for $\theta_{max} = 825°C$ and 910°C, respectively, for a $3\frac{5}{8}$ in. long barrel section. Disregarding irregularities of detail, these traces are seen to conform well with the theoretical curves; in fact, it is almost surprising that in spite of the simplifications made for the computation of the theoretical curves, such satisfactory agreement with experimental work was achieved. The maximum temperatures (θ_{max}) encountered on these and other similar records are found to be between 800°C and 950°C.

Speculating on the causes for the irregularities in the traces, it can be said that the vibrations seen on most traces are probably caused by a longitudinal oscillation of the bullet; careful degreasing of bullet and barrel had the effect of accentuating this vibration as well as of raising the interface temperature. A local increase in

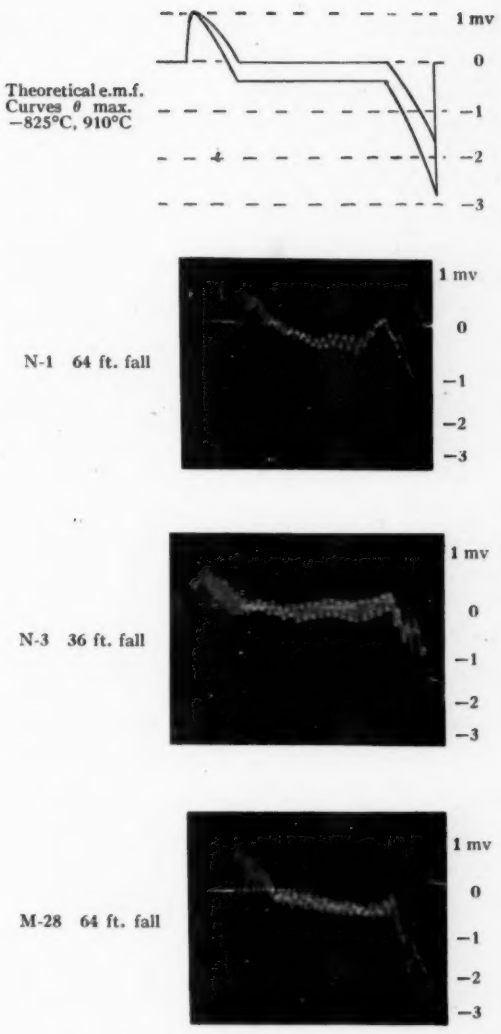


FIG. 6. Interface thermal e.m.f. traces taken with falling weight apparatus. Barrel section No. 4 (0.500″/0.520″)—40-lb. carriage. Timing dot intervals 1/5000 sec. Oiled M2 bullets.

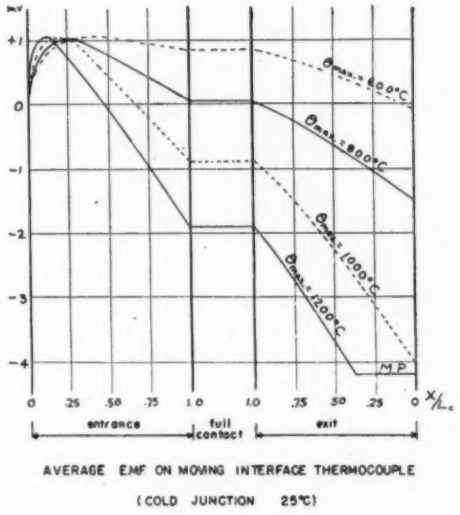


FIG. 5.

frictional pressure causes additional heat development and, consequently, a rise in the interface temperature; it is therefore probable that some of the deviations from theoretical curves are due to variations of pressure. The passage of the crimped portion of the bullet jacket might be expected to be noticeable in the traces, and, indeed, the secondary peak in most exit traces may be attributed to it. Finally, since interface temperatures go up with an increase in the velocity of the bullet, the traces for 64-foot drops are seen to correspond to slightly higher temperatures than those for 32-foot drops. In the

first approximation, the interface temperature rises with the square root of the velocity,[3] but, at higher temperatures the coefficient of friction decreases, and the interface temperature eventually levels off. In the case of drops with low initial kinetic energy, where the bullet velocity decreases appreciably during the travel through the barrel section, the e.m.f. traces observed were found to be distorted in accordance with the expected decline in the interface temperatures.

Figure 7 shows four records which were taken on a 2-in. long barrel section. These records are quite similar to those seen in Fig. 6, but, since they were taken on a much shorter barrel section, the flat central portion of the traces is correspondingly shortened. While the general agreement among traces which were taken under similar conditions is quite satisfactory, it is not surprising to find a number of minor variations between individual traces, due to small changes in local conditions between individual rounds.

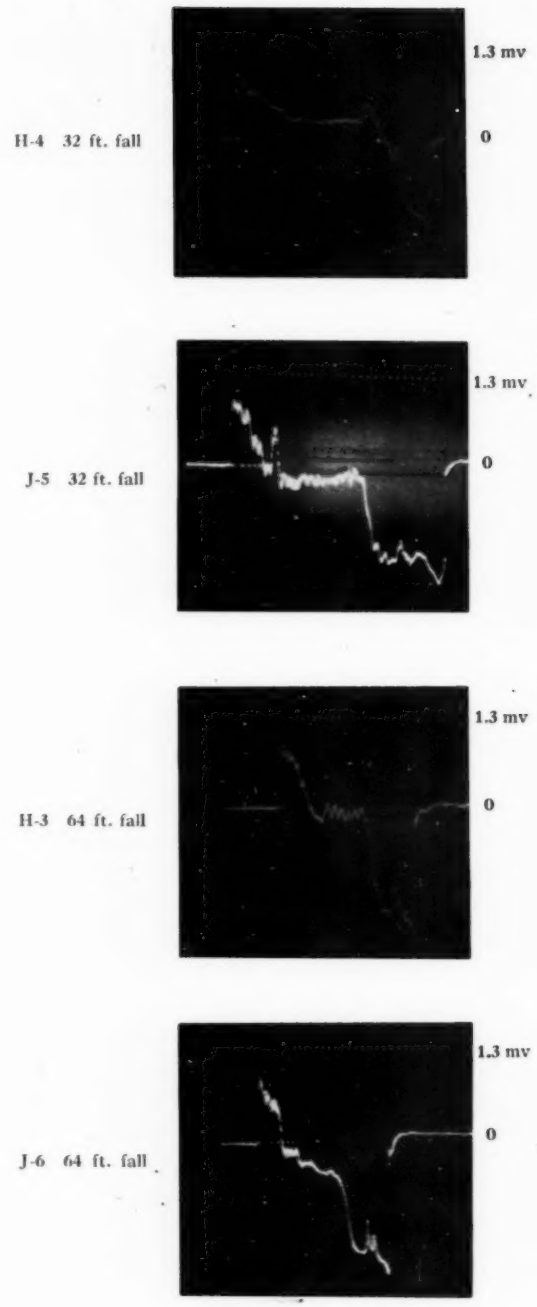


FIG. 7. Interface thermal e.m.f. traces taken with falling weight apparatus. Barrel section No. 3 (0.500″/0.520″)—oiled M2 bullets. Timing dot intervals 1/5000 sec.—40-lb carriage.

INTERFACE THERMAL E.M.F. TRACES FROM FIRED BULLETS

After the work with the falling weight apparatus had given encouraging results, an attempt was made to obtain records of the frictional interface e.m.f. for actually fired bullets. The problem of keeping an insulated wire attached to the tip of the moving bullet was considerably simplified when the barrels were cut down to only 6 in. of rifled length. Experiments undertaken with these short barrels gave quite satisfactory results.

Since it was found desirable to correlate the recorded instantaneous thermal e.m.f. with the position of the bullet, a simple apparatus was devised to record the time when the bullet passed through certain points. A number of hardened steel wires were placed at fixed distances in an insulated holder over the muzzle end of the shortened barrel. A disk mounted on a "harpoon" which extended from the front end of the bullet, touched and broke these steel wires in succession, as the bullet moved forward, and thereby closed and opened a circuit which applied a fixed voltage to the CRO. A leaf-spring contact, mounted at the rear end of the barrel receiver was automatically closed by the motion of the firing mechanism and thereby made to initiate a single sweep on the CRO; the closing and opening of the circuit through the bullet harpoon and the

[3] J. C. Jaeger, reference 1, Eq. (33).

steel wires, during the sweep, was used to obtain position-time records.

The electrical set-up for taking frictional interface e.m.f. traces during the firing was the same as that for the falling weight apparatus, with the above-mentioned contact on the barrel receiver taking the place of the spikes on the carriage and the twin mercury pools as switching device for the initiation of a single sweep on the CRO.

Traces obtained with a fired bullet are essentially similar to those from the falling weight apparatus, however, because of the much higher velocity of a fired bullet, fusing is found to occur over a considerable portion of the frictional area; the simple parabolic temperature distribution over the length of the area of contact, which is valid with a constant coefficient of friction, does therefore not hold for this case. Furthermore, since the CRO traces give records of e.m.f. *vs.* time rather than *vs.* travel, the traces obtained have to be redrawn with the use of position-time records, if records of e.m.f. *vs.* travel are desired.

THEORY OF CALORIMETRIC MEASUREMENTS IN BARREL SECTIONS

In addition to the recording of interface thermal e.m.f.'s, the experimental set-up was well suited and easily adapted for calorimetric measurements of the amount of frictional heat transferred to the barrel wall. It appeared of interest to carry out such measurements and to compare the results obtained with heat input values which were calculated from bullet-bore interface temperature determinations.

A heat input H per unit area of bore wall, in a given unit length of barrel (of annular cross section), causes a temperature rise θ_e throughout that length amounting to, after temperature equilibrium has been reached,

$$\theta_e = \frac{H}{bc\rho}\frac{1}{1+b/d}, \qquad (2)$$

where b is the thickness of the barrel wall in the section, d the bore diameter, c the specific heat of the barrel material, and ρ its density.

The time required to reach thermal equilibrium throughout the barrel section varies approximately with the square of the thickness of the barrel wall,[4] and, *e.g.*, for a wall thickness of $\frac{1}{4}$ in., it is of the order of 2 sec.

While the reading of the "barrel calorimeter" is not affected by a continuous variation in the heat input over the length of the barrel, a discontinuity in the amount of the heat input, or a change in the cross section of the barrel, or any other "end effect" does influence the temperature in the cross section in which the measurement is taken. Thus, a temperature difference δ_e, which is maintained at a distance x, causes a temperature difference δ in a cross section of thermal diffusivity a^2, after a time t, according to[5]

$$\delta/\delta_e = 1 - \mathrm{erf}(x/2at^{\frac{1}{2}}). \qquad (3)$$

Therefore, although the calorimetric measurement cannot be made before radial temperature equilibrium has been attained, it is essential that the reading or recording be completed before any temperature difference due to axial flow becomes appreciable. *E.g.*, a value of $\delta/\delta_e = 0.0025$ is reached in steel after about 25 sec. with $x = 3$ in.; with $x = 1$ in., the same value is reached after slightly less than 3 sec.

Depending on the time required to carry out the temperature measurement (whether done manually or by some automatic recording equipment) the distance x has to have a definite minimum value. *E.g.*, for heat input determinations in caliber 0.50 firings, the distance of a thermocouple from the end of the "calorimetric section" of reduced outside diameter could be kept as short as 1 in., when measurements were made with a Leeds and Northrup Speedomax A (electronic potentiometer recorder); this distance allowed a time of about 1 sec. between the establishment of radial temperature equilibrium and the beginning of a perceptible temperature decay in the section, caused by heat flow to the ends of the "calorimetric section."

The amount of heat received by any point on

---

[4] $\theta/\theta_e = 1 + \sum_{n=1}^{\infty} (-1)^n \exp(-n^2\pi^2 a^2 t/b^2)$.

See H. S. Carslaw, *Mathematical Theory of the Conduction of Heat in Solids* (The Macmillan Company, London, 1921), second edition, Chap. IX.

[5] H. S. Carslaw, reference 4, Chap. III.

the bore during passage of a bullet is given by[6]

$$H=\left(\frac{kc\rho}{\pi}\right)^{\frac{1}{2}}\int_0^{\tau}\frac{\theta(t)}{(\tau-t)^{\frac{1}{2}}}dt, \tag{4}$$

where H is the heat input per unit area of bore, k the thermal conductivity, c the specific heat, ρ the density of the barrel material, θ the temperature rise of the bore as a function of time t, and τ the total time of contact of the point concerned.

Assuming that it is permissible to approximate the temperature distribution over the contact area of the bullet with the bore by Eq. (1), which is valid for a "slider" moving under steady-state conditions, the temperature of a fixed point on the bore, during passage of the bullet, is given by:

$$\theta(t)=\theta_{\max}(t/\tau)^{\frac{1}{2}}. \tag{5}$$

Substituting (5) into (4) and integrating, one obtains finally:

$$\theta_{\max}=\frac{2}{\pi^{\frac{1}{2}}}\frac{H}{\tau^{\frac{1}{2}}(kc\rho)^{\frac{1}{2}}}. \tag{6}$$

RESULTS OF CALORIMETRIC MEASUREMENTS

For heat input measurements in the falling weight apparatus, the "calorimetric section" had a length of $3\frac{5}{8}$ in., with the couple peened in its center; no thermal insulation at the ends of the section was required. A special Leeds and Northrup mirror galvanometer recorder was used as recording equipment.

The frictional heat input values obtained by the "calorimetric section" method scatter rather widely, averaging 2.85 cal./cm² and 3.65 cal./cm² for drops from 64 feet and 32 feet height, respectively, when referred to the whole bore surface; when only the area of actual contact (assumed to consist of the top and about one-half of the sides of the lands) is taken into account, these values rise to 8.7 cal./cm² and 11.0 cal./cm², respectively, in a 0.500″/0.520″ rifled barrel section, at about $1\frac{3}{4}$ in. beyond the origin of rifling.

When these heat input values are substituted in Eq. (6), with $(kc\rho)^{\frac{1}{2}}=0.31$ cal./°C cm² sec.$^{\frac{1}{2}}$ (for steel) and a length of contact $L_0=2.0$ cm, values of $\theta_{\max}$ between 950°C and 1000°C are obtained for either case (with bullet velocities of about 60 ft./sec. and 40 ft./sec., respectively). These temperatures are somewhat higher than those calculated from sliding thermocouple records (800°C to 950°C); however, in view of the simplifying assumptions made for the calculations and of the various uncertainties involved, the agreement can be considered quite satisfactory. Furthermore, the errors in the temperatures calculated are in the direction to be expected; while Eq. (4) is based on the assumption of linear heat flow from the bore surface into the barrel, heat actually is generated over only a fraction of the bore surface (top and about one-half of sides of lands) in barrels used for this work, and it spreads from there two-dimensionally, with the effect that a lower temperature than with strictly linear flow is required at the contact area in order to produce a given heat input; this effect increases rapidly with duration of the time of contact, and is therefore more pronounced for 32 foot drops than for 64 foot drops.

Concluding, it may be said that frictional temperature determinations by means of the sliding thermocouple method gave satisfactory results in an application where no other known methods of temperature measurement could be used.

[6] See H. S. Carslaw, reference 4, p. 47, and note that $H\equiv\int_0^{\infty}c\rho\theta_x dx$, where θ_x denotes the temperature at any point x.

Carbide Structures in Carburized Thoriated-Tungsten Filaments

C. W. HORSTING
Tube Department, Radio Corporation of America, Harrison, New Jersey
(Received April 16, 1946)

A wide variety of carbide structures is described, as found in the surface layer of carburized thoriated-tungsten filaments. Their origin is traced back to carburizing conditions and subsequent processing during tube making. A frequently occurring laminated carbide structure is found to contain less carbon than W_2C. The thyratron relay method of carburizing control is critically reviewed. Abnormal filament current in manufactured tubes is explained as due to surface conditions which cause changes in thermal emissivity.

THE application of carburized tungsten-filament cathodes is a well-known practice in radio tube manufacture, especially in the power tube field where high electron emission is required at high voltages. By the use of this method, longer filament life and higher operating temperatures are possible because, it is claimed, the emission-causing thorium layer at the filament surface adheres more firmly to the tungsten-carbide surface than to the pure tungsten surface.[1]

During the production of power tubes, the tungsten filaments are partially converted into tungsten carbide by heating them electrically in a hydrocarbon atmosphere to a temperature of about 2200°C. The hydrocarbon is decomposed at the hot filament surface to produce free carbon which diffuses into the tungsten, forming tungsten carbide. The degree of carburization is limited by the fact that the mechanical strength of the filament decreases rapidly with increasing carbon penetration. In practice, therefore, the carburizing reaction has to be stopped accurately at a predetermined penetration.

For this purpose, and chiefly because of the high speed of the carburizing reaction, a thyratron-controlled relay is often employed to cut off the heating current and, thus, stop the reaction after the desired amount of carbide has been formed. The fact that tungsten carbide has a higher electrical resistance than tungsten is utilized to determine the action of the relay. Since carburization is allowed to take place at a constant heating current, a rising voltage appears across the filament terminals as the resistance of the filament increases. The thyratron-operated relay is adjusted to function at a desired point on the rising voltage curve. Ideally, this point corresponds to a constant degree of carburization.

However, under certain circumstances hitherto unknown, filaments were produced which showed abnormal properties. The main difficulty encountered was too large a spread of the cold-resistance and the hot-resistance values of the finished tubes and inconsistent correlation between cold and hot resistance values.

In order to determine the reasons for these variations, the following investigation was carried out. At the start of the investigation, the main questions could be briefly stated as follows: Why does the theoretically well-founded method of thyratron control fail to produce tubes with constant filament properties? What are the circumstances under which the thyratron control fails? Is the resistance-control method unsuitable? What secondary phenomena may be causing the discrepancies?

In order that answers to these questions could be obtained, filaments having abnormal characteristics were collected and analyzed metallographically as to their microstructure. A wide variety of structures was found. Partial cross sections of two of these filaments are reproduced in Figs. 1 and 2. Figure 1 shows a filament which was found to be normal, whereas Fig. 2 shows a filament with normal cold resistance, but with low hot resistance, and a resultant filament current 6 to 7 percent above the value normally occurring at the specified operating voltage.

It has been well-known that different structures and forms of tungsten carbide can be produced under different circumstances. W. P.

[1] L. R. Koller, *The Physics of Electron Tubes* (McGraw-Hill Book Company, Inc., New York, 1937).

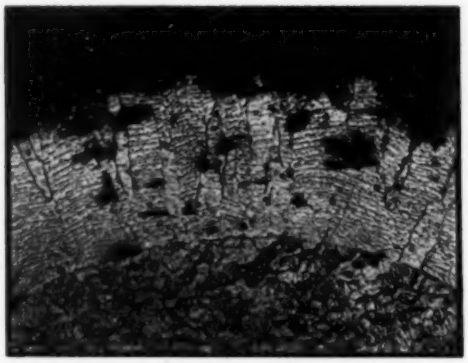


FIG. 1. Normal carbide structure. Magnification 250×.

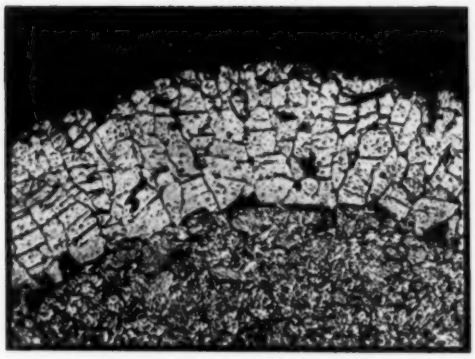


FIG. 2. Abnormal carbide structure. Magnification 250×.

Sykes,[2] K. Becker,[3,4] B. T. Barnes,[5] M. R. Andrews,[6] and others have reported on a variety of carbide structures, but the relation between the different carbide formations present in carburized tungsten filaments and their origination from the carburizing process had hitherto not been established. Consequently, efforts were made to reproduce under laboratory conditions the formations found in filaments and, thus, to determine the factors that produced them.

For that purpose the apparatus shown in Fig. 3 was constructed. With this apparatus, a U-shaped filament can be electrically heated in a hydrogen atmosphere to which can be added a controlled amount of hydrocarbon vapor by means of a specially devised carburetor. The current through the filament is controlled by means of a Variac auto-transformer. The carburizing time is regulated by a time switch which operates a relay switch in the main circuit. Such filaments carburized with this apparatus gave accurate reproduction of carbide formation and of electrical resistance.

One of the first filaments produced with this apparatus had a carbide structure as shown in Fig. 4. About half of the carbide shell consists of a laminated carbide formation similar to the one found in the normal filament. The other half is of a massive structure. By changing the hydrocarbon content of the hydrogen gas, it was possible to change the proportions of the two formations in the carbide shell at will. By lowering the hydrocarbon content, more of the laminated structure appeared and, simultaneously, more time was required for a certain depth of penetration of the carbide, whereas increasing the hydrocarbon content had the opposite effect. Figures 5a, b, and c show three different samples which were carburized to the same depth but with increasing hydrocarbon concentrations.

Because it is not possible to reproduce photomicrographs of all the cases found, a graphical reproduction of the whole range of structures obtained with a wide variation of hydrocarbon concentrations is given in Fig. 6. As a key to the interpretation of Fig. 6, a graphic picture corresponding with the photograph shown in Fig. 4 is given in Fig. 7. It will be explained later on how the concentration-penetration diagrams were arrived at. In Fig. 6, the diagrams of carbon

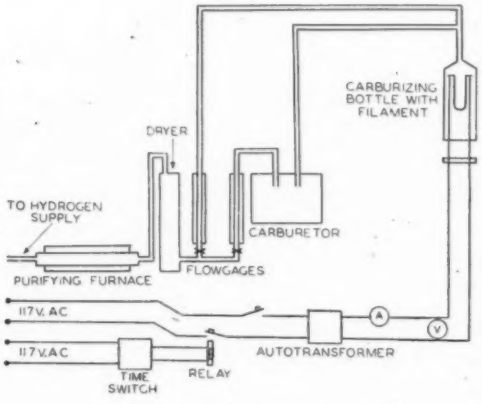


FIG. 3. Carburizing apparatus.

[2] W. P. Sykes, *A Study of the Tungsten-Carbon System.* Paper presented at the 12th annual convention of the Am. Soc. for Steel Treating, Chicago, 1930.

[3] K. Becker, "Ueber das system Wolfram-Kohlenstoff," Zeits. f. Metallkunde **20**, 437–41 (1928).

[4] K. Becker, "Die Konstitution der Wolfram-Karbide," Zeits. f. Electrochemie **34**, 640–642 (1928).

[5] B. T. Barnes, "Properties of carbonized tungsten," J. Phys. Chem. **33**, 688–691 (1929).

[6] M. R. Andrews, J. Phys. Chem. **27**, 275 (1923).

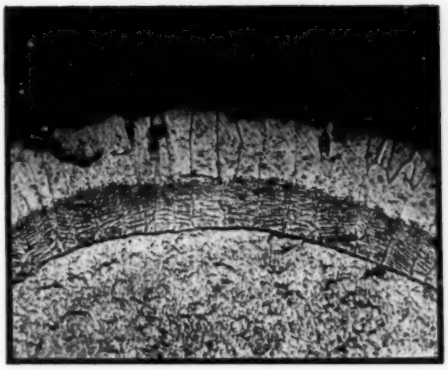
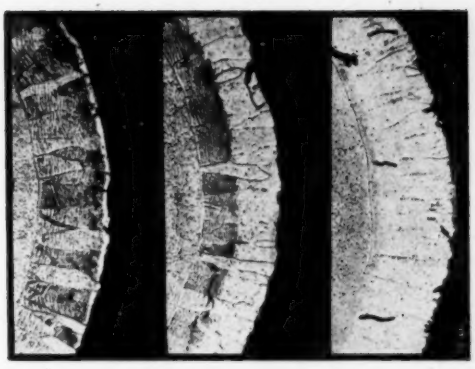


FIG. 4. Carbide shell containing solid as well as laminated carbide structure. Magnification 250×.



a b c

FIG. 5. a, b, and c. Carbide shells produced by carburization to the same depth of penetration, but with increasing hydrocarbon content of the hydrogen atmosphere. Magnification 175×.

concentration *versus* penetration correspond to the structures shown at the left.

Figure 6a shows the effect produced by carburizing a filament in an atmosphere rich in hydrocarbon. A carbon deposit covers the surface, and two massive carbide phases are separately visible. The inner one is almost pure W_2C, whereas the outer one is WC. The concentration-penetration diagram for this sample is shown on the right.

Figure 6b is the same but without the carbon deposit.

Lowering the hydrocarbon content somewhat more prevents the formation of the second solid phase, the result of which is shown in Fig. 6c.

In Fig. 6d, we see the appearance of the laminated phase which was found in the normal filament; Figs. 6e–6h show the effects of reducing the hydrocarbon content further. The massive structure decreases until only widely spaced laminated material is formed at the very lowest hydrocarbon concentration.

The fact that the laminated structure is connected with a low hydrocarbon concentration, and the massive structure with a higher one, suggests that the laminated material contains less carbon than the massive W_2C phase.

In addition, we know that where only laminated material is formed, the layers become wider as the hydrocarbon content of the carburizing atmosphere decreases.

Combination of these two facts leads to the conclusion that the change in character which we observed in the laminated phase in Fig. 4 actually shows us a concentration gradient of decreasing carbon content from the massive phase toward the core. The gradual transition between the massive W_2C phase and the laminated phase

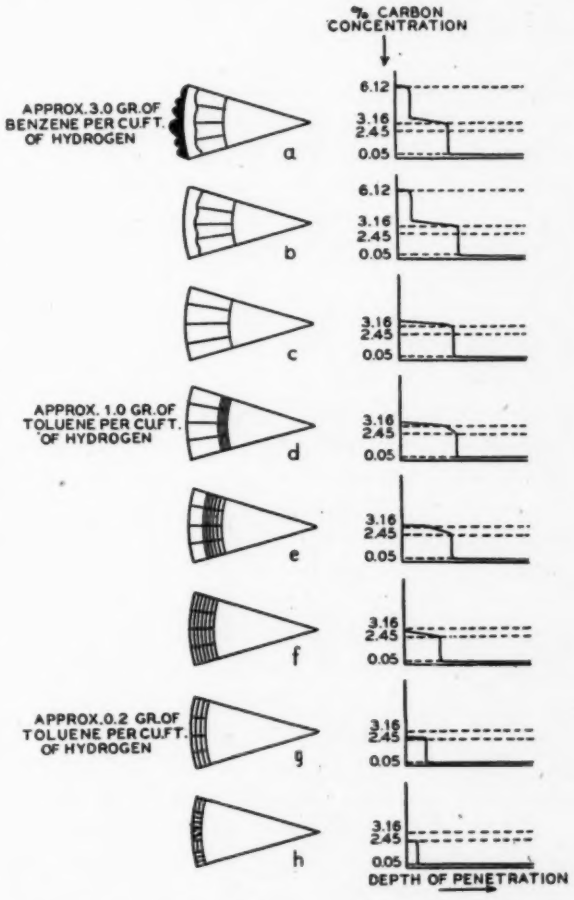


FIG. 6. Carbide formations produced by carburization with constant carburizing time and constant starting temperature, but with decreasing hydrocarbon content of the hydrogen atmosphere.

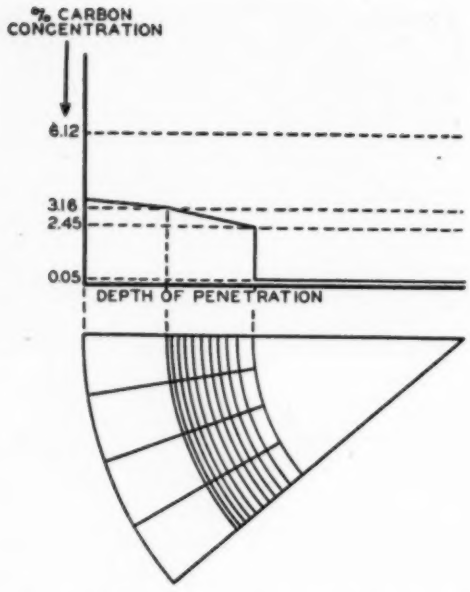


FIG. 7. Key to the interpretation of Figs. 6 and 11.

suggests that the carbon deposit at this side of the gradient starts at 3.16 percent. The sharp line of demarcation between the laminated phase and the core suggests that the gradient does not end in a zero carbon concentration, but abruptly drops from a certain value to zero or some very small value.

In order to test the idea of a gradient, it was reasoned that a gradient of reverse character should be produced if carbon were removed from a previously carburized filament. This was done by heating such a filament in moist hydrogen. Figure 8b shows the results obtained. Figure 8a shows the original structure. The partial removal of carbon from the filament produced a laminated layer near the surface of the filament. It is clearly visible in the cross section that the laminations are widest near the surface where even a small decarburized layer is visible. The gradient on the inside of the remaining W_2C section is, of course, of decreasing carbon content toward the core. The opposite character of the inward and outward gradients is clearly visible in Fig. 8b.

The explanation of the concentration-penetration diagrams in Fig. 6 is based on the following considerations and experiments.

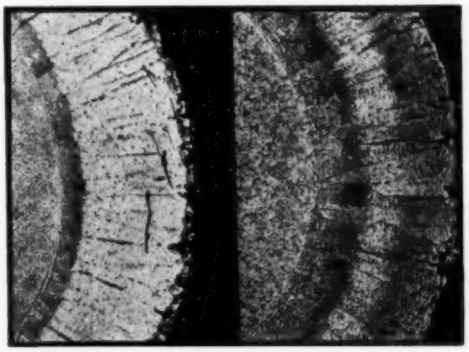
If we briefly review the theory of diffusion in solids, it does not seem to be difficult to explain the structures which are reproduced in Fig. 6.

If diffusion occurs in a binary system, and the components show mutual solubility in all proportions, one can generally expect a smooth concentration-penetration curve showing a gradual concentration decrease from the surface inward and ending in a zero concentration of the penetrating element at some distance from the surface.

However, in case limited solubility occurs, an area of co-existent phases separates the solubility fields in the phase diagram. If diffusion takes place in such a system, a sharp break in the concentration-penetration curve can be expected. The concentration drop at this break corresponds to the difference in solute content of the co-existent phases at the temperature involved, while the absolute concentrations on either side of the break are equal to the solute contents of the two co-existing phases.

In the somewhat more complicated case where compounds are formed between the components of the binary system, several areas of co-existence may occur, which are separated by more or less wide areas of solid solubility. Corresponding to each area of co-existence, a sudden break in the concentration-penetration curve must be expected. Each area of solid solubility corresponds to a gradient of gradually changing concentration.

Let us now consult the available data on the system tungsten-carbon and see what has to be expected in view of the above (see Fig. 9). Two compounds occur between tungsten and carbon, W_2C and WC, the first containing 3.16 percent carbon, the other 6.12 percent carbon. Starting



a b

FIG. 8. Carbide shells. a, before, b, after, partial decarburization in wet hydrogen. Magnification 175×.

from the tungsten side of the diagram and only considering solids, we find successively:

1. A very narrow field of solid solubility of carbon in tungsten (W. P. Sykes reports a maximum solubility of 0.05 percent carbon in tungsten at 2400°C).
2. A co-existent phase area.
3. A narrow field of solid solubility around the stoichiometric W_2C composition.
4. A co-existent phase area.
5. The compound WC with no solid solubility field.
6. An area of co-existence between WC and graphitic carbon.

If we apply the afore-mentioned theory to these data, we can expect the shape of the concentration-penetration curve for carbon diffusing into tungsten to be as indicated in Fig. 10. It is obvious, of course, that the complete sequence of concentration changes from the surface inward can occur only if sufficient carbon is available at the surface, and that the relative amounts of the different formations will depend upon the diffusion velocities within those formations and the degree of equilibrium reached during the diffusion reaction.

The concentration gradient within the solid solution range is, for simplicity's sake, indicated as a straight line. It will generally be a curved line which may be convex or concave to a larger or lesser degree depending on carburizing conditions. With this diagram as a basis, it is comparatively simple to construct the concentration-penetration diagrams, which are given in Fig. 6. If, on the basis of the above analysis, we reconsider the laminated gradient, it becomes clear that this material must be identified with the solid solution range of tungsten in W_2C. In order to explain its laminated character we have to invoke Sykes' claim that the solubility of tungsten in W_2C decreases with temperature and that the tungsten-rich phase precipitates out on cooling from the diffusion temperature. Heavier layers are formed where more tungsten was in solution, while the wider spacing of the layers occurring simultaneously, may be attributed to the higher precipitation temperature. That no precipitation seems to occur where excess carbon is dissolved in W_2C would indicate

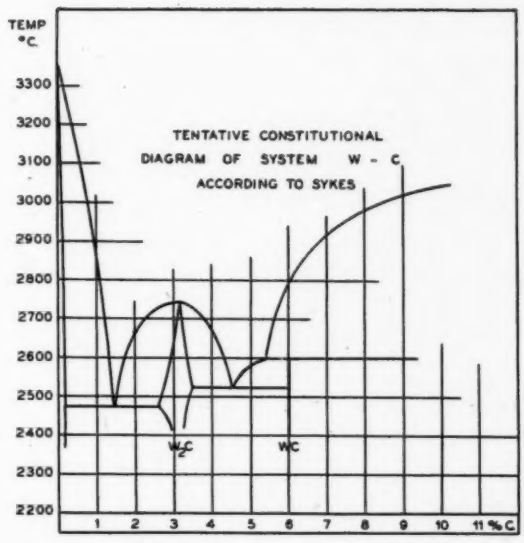


FIG. 9. Tentative constitutional diagram of tungsten carbon (used with permission).

that the solubility of carbon in W_2C does not decrease with temperature or that it is small and shows no pronounced effect on precipitation.

The existence of the carbon concentration gradient, not only in the laminated region, but also in the massive carbide, was made plausible by a series of micro hardness tests. Successive tests, starting from the center of the wire outward toward the surface, were made with a Tukon hardness tester on a sample like the one shown in Fig. 4 containing both the massive and the laminated material. The core was found to be of uniform hardness. A sudden sharp increase in hardness occurred at the demarcation line between core and carbide shell followed by increasing hardness throughout the laminated zone. A maximum hardness was found just inside the massive zone, while the hardness again decreased from there to the surface. The increase in hardness throughout the laminated part of the shell confirms the gradient in this part; the relative amount of W_2C increases. But the hardness distribution in the massive material also indicates an increasing carbon content toward the surface, for it is known that W_2C attains its maximum hardness at the stoichiometric composition. Excess carbon, however, again lowers the hardness.

Finally, a direct carbon analysis of the laminated material was made in order to establish definitely its carbon content. At the same time

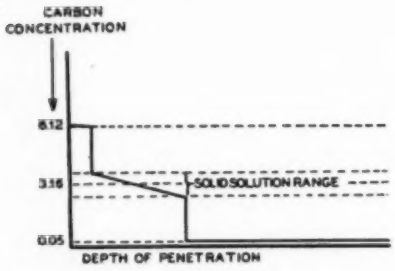


FIG. 10. Theoretical concentration-penetration curve for the diffusion of carbon into tungsten.

an attempt was made to determine the solubility limit of tungsten in W_2C or, what amounts to the same thing, the carbon concentration at the carbide-side of the sharp interface between the carbide shell and the core. If carburization is carried out at the lowest hydrocarbon concentration which will allow formation of laminated material, and the temperature is sufficiently high to guarantee a diffusion reaction which is essentially "carbon-limited," it is safe to assume that equilibrium conditions are closely approached and that whatever gradient is formed is so small as to be negligible. The concentration-penetration diagram of this laminated material is indicated in Fig. 6h, and the whole carbide shell is of a widely spaced laminated character which does not show any evidence of a gradient. This material was produced and analyzed.

The carbon analyses were carried out with the aid of a high vacuum carbon analysis system which was built on the principle first described by Yensen,[7] later improved by Ziegler,[8] and redesigned by Wooten and Guldner[9] and others. The process, briefly described, utilizes a sample weighing about 10 milligrams which is heated in a vacuum to 1000°C and then burned in an excess of purified oxygen. After the CO_2 formed is frozen out in a liquid-air trap, the excess oxygen is removed by means of a vacuum pump. The CO_2 is then permitted to expand into a known volume. The CO_2 pressure at room temperature then furnishes the data required for a calculation of the carbon content of the sample under analysis.

At the same time, cross sections were made of the immediately neighboring part of the filament; the relative areas of the carbide shell and the tungsten core were measured by the use of a planimeter on the tracing of the enlarged cross-section image. This information made it possible to calculate the carbon content in the shell.

Results are given in Table I.

From the results obtained it can be seen that the carbon content is lower than 3.16 percent, as expected and, furthermore, that the equilibrium condition mentioned above was closely approached. This is borne out by the fact that the values obtained are fairly constant, even though the samples were carburized to different depths. We have assumed the average value of 2.45 percent carbon to be the one corresponding to the solubility limit of tungsten in W_2C, and the diagrams in Fig. 6 are drawn accordingly.

If we compare Sykes' tentative curves (Fig.9) with the result of our experiment, it seems that the value of 2.45 percent is somewhat low. However, the diffusion method is one of the most sensitive ones for the investigation of solid solubility limits, and Sykes' constitution diagram does not claim great accuracy in this region.

During the study of the decarburization process, it was found that the conditions shown in Fig. 8b could be produced only by high speed decarburization, that is, in wet hydrogen where oxygen is present to react with the carbon and remove it quickly from the filament surface as CO or CO_2 gas. In dry hydrogen, decarburization is much slower, and the massive W_2C layer almost entirely disappears as a result of the carbon penetrating inward before a visible gradient toward the outside is established. A general survey of the stages of decarburization reactions is given in Fig. 11.

Decarburization in high vacuum is still slower than in pure hydrogen and needs a higher temperature. In high vacuum only a very limited amount of gas is present for the carbon to react with and simple evaporation may become necessary. Furthermore, it was found that the diffusion velocity of carbon in tungsten is considerably lower in a vacuum than in the presence of hydrogen. However, the same pattern of de-

[7] T. D. Yensen, "Carbon in iron," Trans. Am. Electrochem. Soc. **37**, 227–245 (1920).

[8] Nicholas A. Ziegler, "Recent developments in the analysis of carbon in iron and iron alloys," Trans. Am. Electrochem. Soc. **56**, 231–238 (1929).

[9] L. A. Wooten and W. G. Guldner, "Determination of carbon in low carbon iron and steel," Ind. Eng. Chem. **14**, No. 10, 835 (1942).

carburization stages can generally be observed. Cross sections of filaments decarburized in a vacuum by high temperature flashing showed decarburized surface layers.

Referring to the structure of the normal filament as reproduced in Fig. 1, it can now be seen that this type of carbide formation was probably produced by decarburization. The hydrocarbon content of the hydrogen gas used during carburization was found to be too high for direct formation of laminated material. It follows, therefore, that decarburization must have taken place during the exhaust cycle when the filament was heated to a high temperature. In practice the duration of this cycle was found to be determined by the speed with which a certain degree of vacuum was obtained. This means that the filament was heated in the presence of gas and that the time varied with circumstances determining the degassing speed of the tube parts. Varying degrees of decarburization can easily be explained this way. If, in addition to that, the carbide formation during carburization does not produce constant results, as was found caused by improper dosing of the hydrocarbon in the hydrogen, it follows that a wide variety of carbide structures must be expected.

Varying amounts of carbide, of course, would cause differences in the values of cold resistance and of filament current (hot resistance). But why should two filaments having the same values of cold resistance have different values of filament current? This phenomenon was theoretically explained as follows: The cold resistance of a filament depends upon the amount of W_2C formed, but it does not matter where in the filament this W_2C is present. It may be on the surface as in the case of a freshly carburized filament; or it may be under the surface as in the case of a decarburized filament. Therefore, two filaments can have the same cold resistance, but different surface conditions.

TABLE I.

Weight of sample in grams	% carbon in sample	% carbon in shell	
0.0200	0.350	2.45	
0.0161	0.403	2.31	
0.0136	0.538	2.45	Average
0.0141	0.534	2.36	2.45
0.0155	0.508	2.69	
0.0258	0.400	2.45	

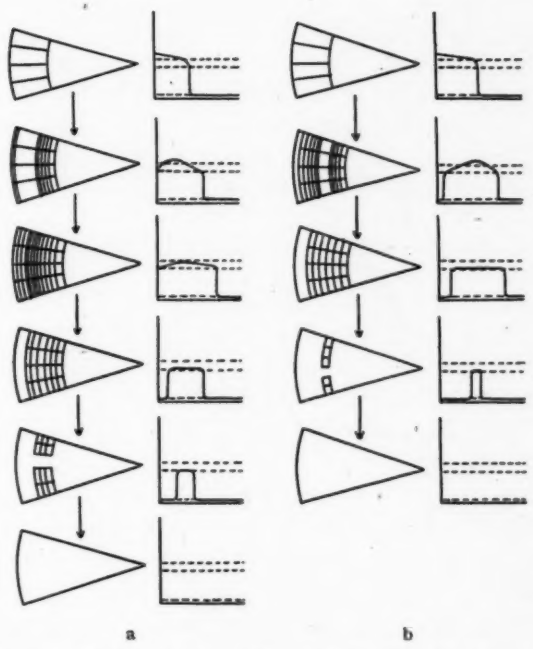


FIG. 11. Series showing effects of progressive decarburization: a, in dry hydrogen; b, in wet hydrogen.

When, however, these filaments are both put on the same operating voltage, as is done during tube operation, the one with the tungsten surface, having a lower heat emissivity, will have a higher temperature, a higher hot resistance, and a lower filament current than the one with the tungsten carbide surface, which has a higher heat emissivity.

The fact that surfaces intermediary between almost pure tungsten and total tungsten carbide as extremes may occur, due to decarburization on exhaust, opens the possibility of a range of hot resistances, corresponding to each cold resistance value. Thus, the cold resistance-*versus*-filament current diagram becomes an area instead of a line.

For confirmation of this theory, special tubes were built containing two identical tungsten filaments having the same initial cold resistance. One of the filaments was carburized under circumstances which guaranteed the formation of the massive W_2C phase. The other filament was overcarburized and, subsequently, decarburized in wet hydrogen to exactly the same cold re-

TABLE II.

Carbide-surface filament				Tungsten-surface filament			
V	*A*	*VA*	*T*°C	*V*	*A*	*VA*	*T*°C
7.5	2.08	15.6	1528	7.5	1.9	14.25	1660
10.0	2.37	23.7	1725	10.0	2.21	22.1	1815
12.5	2.63	32.9	1904	12.5	2.54	31.75	1960
14.3	2.8	40.0	1970	14.3	2.70	38.6	2000
15.3	2.9	44.4	2000				

sistance as the first filament. When these two filaments were operated in series, the decarburized filament showed a visibly higher temperature.

Measurements made on the separate filaments are given in Table II.

These data show that the purposely decarburized filament has the lowest filament current and the highest temperature at a fixed operating voltage. In this case the difference in filament current values at 2000°C amounts to about 4 percent.

If we were to operate the filaments at 14.3 volts, the temperature difference is 30°C. Furthermore, these data agree with Barnes' results,[5] *i.e.* that the total emissivity of tungsten carbide is higher than that of tungsten for the same temperature of operation.

This is an appropriate place to return to our first question as to why the thyratron control sometimes failed to produce filaments of constant cold resistance, because the answer was also found to be differences in surface conditions. A filament with a surface having total emissivity higher than average will produce a lower than normal temperature for normal input. Translated into terms of hot resistance this means that a lower temperature corresponds to the hot resistance at which the thyratron relay is set to act. This in turn requires that overcarburization must take place if the necessary hot resistance is to be reached. A filament of higher than normal cold resistance is the result.

Two cases of abnormally high emissivity of filaments were encountered. The first one occurred when helical filaments were oxidized heavily during the forming operation and then were reduced by flashing them in a hydrogen atmosphere. This procedure produced a very rough surface. In the second case, the hydrocarbon content of the carburizing atmosphere was so great that a carbon deposit was formed on the surface of the filament.

From the above investigations, the following conclusions can be drawn:

1. Thyratron control for the carburization of thoriated-tungsten filaments is an excellent method, provided the filaments to be carburized have uniform surface conditions, and provided a strict control of the hydrocarbon content in the hydrogen atmosphere is maintained.

2. The character of the carbide formations obtained during carburization and decarburization can be explained on the basis of the theory of diffusion and the characteristics of the carbon-tungsten phase diagram.

3. The amount of decarburization which occurs during tube manufacture has an important effect on the uniformity of the electrical filament characteristics of tubes.

4. The hot-to-cold resistance ratio of a filament is greatly influenced by the surface condition of the filament.

Creep Deflections in Columns

JOSEPH MARIN*
School of Engineering, The Pennsylvania State College, State College, Pennsylvania
(Received August 22, 1946)

The determination of column deflections and column buckling loads has been considered for many years. The available theories, however, do not provide for materials which creep with time under constant loads. For the design of structural members made of these materials, a consideration of creep may be of practical importance. Plastics, concrete, and some metals creep at normal temperatures while other metals creep only at high temperatures and at stress values beyond the yield point. A consideration of creep in the design of some structures appears appropriate in view of the modern developments in plastics and the presence of high stress values which are sometimes beyond the yield stress. This paper gives a rational theory for predicting creep deflections in columns. A special case using this theory is applied to the interpretation of some preliminary tests of an aluminum alloy.

INTRODUCTION

THE usual concept of failure in columns by elastic instability must be modified if the material creeps. For columns made of materials that creep, critical buckling loads are smaller than the elastic buckling loads since the progressive lateral creep deflections must be considered. It is necessary therefore to modify our definition of the critical buckling load and to define the buckling load as that load which will produce a large lateral deflection at the end of a specified time. To determine these loads, a theory will be developed giving the creep deflections in a column in terms of the load, column dimensions, material constants, and time. These theoretical deflections will then be compared with the corresponding experimental values.

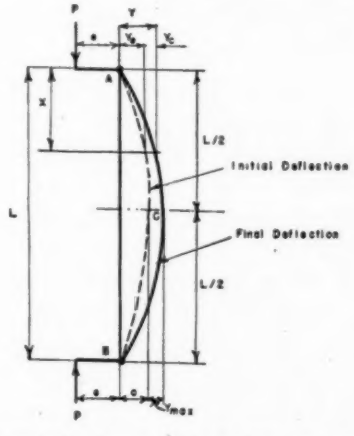


FIG. 1. Pin-ended column subjected to an eccentric load.

* Professor of Engineering Mechanics.

GENERAL THEORY FOR COLUMN DEFLECTIONS

Columns having cross sections with at least one axis of symmetry will be considered. The column AB is free at the upper end A, and at the base B, and is loaded eccentrically with a load P in the plane of symmetry (Fig. 1). Initially, an elastic lateral deflection y_e is produced at a distance x from point A. This initial deflection will include the elastic deflection plus the initial creep deflection as shown in Fig. 2. Actually the column deflects so that AB rather than AC represents the initial elastic and creep deformation and beyond B for a short period of time the deflections are represented by the curve BD. Beyond point D the experimental deflection-time relation DE was found to be linear. It will be assumed in this theory that the actual deflection curve $ABDE$ is replaced by ACE. To determine the lateral deflection $(y=y_e+y_c)$ at a time t, a theory for defining

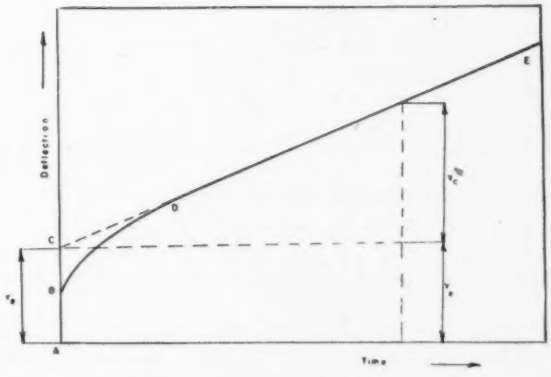


FIG. 2. Deflection-time relation.

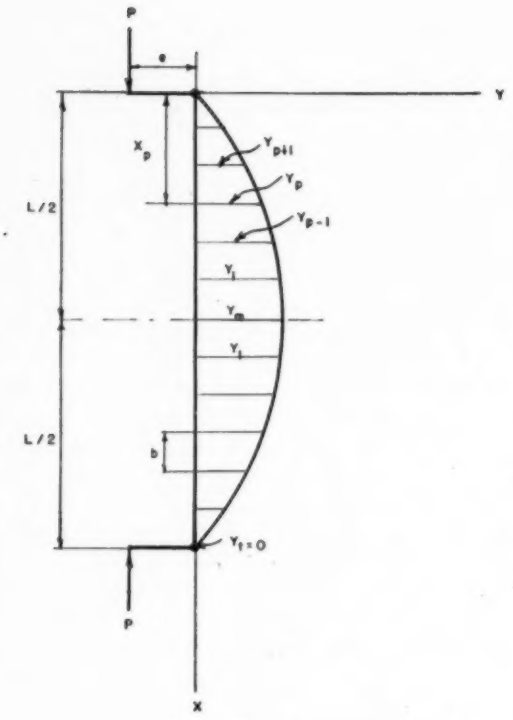


FIG. 3.

creep deflection in bending corresponding to the elastic theory for deflection must first be obtained.

A theory for the creep deflection in bending has been proposed by the writer in the form of a differential equation for the creep deflection y_c in terms of the moment M and time t.[1] It was shown by this theory that

$$(D/t)(d^2y_c/dx^2)=M^n. \tag{1}$$

In Eq. (1), D corresponds to the flexural rigidity EI in the elastic theory. Its value for a rectangular cross section is

$$D=\frac{1}{A}\frac{(2b)^n(h/2)^{2n+1}}{(2+1/n)^n}, \tag{2}$$

where A and n = experimental creep constants,

b = the width of the section,

h = the depth of the section.

Equation (1) is based on the log-log creep stress law $C=AS^n$ where C is the assumed constant rate of creep strain and S is the unit stress.

The moment M in terms of the load P at a time t is, from Fig. 1,

$$M=-P(e+y)=-P(e+y_e+y_c), \tag{a}$$

where e = the eccentricity of the load. Substituting the value of the moment M from Eq. (a) in Eq. (1), the differential equation for the creep deflection y at a distance x is

$$(D/t)(d^2y_c/dx^2)=-P^n(e+y_e+y_c)^n. \tag{3}$$

In this equation x, y_e, and y_c are variables. These three variables can be reduced to the necessary two by expressing the initial deflection y_e in terms of x. For this purpose the shape of the curve AB can be assumed with reasonable accuracy. This assumption is similar to that used in the analysis of buckling problems in the elastic case using the Rayleigh method. For the present problem the shape of the curve AB will be assumed as a sine function. That is, the value of y_e in terms of x will be assumed as

$$y_e=a\sin\frac{\pi x}{L}, \tag{b}$$

where a = the initial deflection at point C in the column and L = the length of the column.

Placing the value of y_e from Eq. (b) in Eq. (3) and noting that $y=y_e+y_c$, the differential equation for the total deflection y as a function of the distance x becomes

$$\frac{d^2y}{dx^2}=-\frac{a\pi^2}{L^2}\sin\frac{\pi x}{L}-\frac{P^nt}{D}(e+y)^n. \tag{4}$$

The values of n in Eq. (4) must be assumed as odd integers since the creep stress law $C=AS^n$ must apply for compressive and tensile values of S in the case of bending where both types of stress exist.[2] That is, Eq. (4) must be solved for values of $n=1, 3, 5\cdots n$, where n is an odd number. A direct solution of Eq. (4) for y in terms of x could not be found. One method of attack using numerical integration is given in Appendix 1. A general method of solution is developed in this paper by using a trial and error

[1] J. Marin, "Creep of aluminum subjected to bending at normal temperature," Proc. A.S.T.M. **40**, 937 (1940); and "Methods of structural analysis in the case of creep," J. West. Soc. Eng. **45**, 301 (Dec. 1940).

[2] J. Marin, "Mechanics of creep for structural analysis," Trans. A.S.C.E. Vol. **108**, 470 (1943).

procedure. This method is similar to the one used by Holzer[3] for the determination of natural frequency of vibration of multi-mass systems. The procedure used is to replace the differential Eq. (4) by an equivalent system of difference equations. To do this, consider the column divided into a number of segments N of length $b=L/N$ (Fig. 3). The ordinates to the deflection curve will be designated by $y_m, y_1 \cdots y_p$, as shown in Fig. 3. The second derivative of y with respect to x, d^2y/dx^2, at any section where $x=x_p$, is given approximately by the equation:

$$\left(\frac{d^2y}{dx^2}\right)_{x=x_p} = \frac{1}{b^2}(y_{p+1}-2y_p+y_{p-1}). \qquad (5)$$

Placing the value of d^2y/dx^2 from Eq. (4) in Eq. (5) and solving for y_{p+1},

$$y_{p+1}=2y_p-y_{p-1}-F_p+Q(e+y_p)^n, \qquad (6)$$

where

$$F_p=(b^2a\pi^2/L^2)\sin(\pi x_p/L), \qquad (7)$$

and

$$Q=b^2P^nt/D. \qquad (8)$$

Denoting the deflection at the mid-section of the column by y_m and at the end by y_t and writing Eq. (6) for each section successively,

$$\left.\begin{aligned} y_1&=y_m-\tfrac{1}{2}[F_m+Q(e+y_m)^n],\\ y_2&=2y_1-y_m-[F_1+Q(e+y_1)^n],\\ y_3&=2y_2-y_1-[F_2+Q(e+y_2)^n],\\ y_{p+1}&=2y_p-y_{p-1}-[F_p+Q(e+y_p)^n],\\ y_t&=2y_{t-1}-y_{t-2}-[F_{t-1}+Q(e+y_{t-1})^n], \end{aligned}\right\} \qquad (9)$$

where

$$F_m=\frac{b^2a\pi^2}{L^2}\sin\frac{\pi x_m}{L}, \quad F_1=\frac{b^2a\pi^2}{L^2}\sin\frac{\pi x_1}{L}. \qquad (10)$$

The method of applying Eqs. (9) to determine column deflections for a given load P at a time t is as follows:

1. From the column dimensions, material constants A and n, time t, and load P, determine the value of Q by Eq. (8).

2. Using the value of the elastic deflection for the approximate value of a and for different values of x determine the values of $F_1, F_2 \cdots F_p$.

3. Assume a value for y_m and, by Eqs. (9), determine successively values of y_1 to y_t. The value of y_t should be zero if the correct value of y_m was assumed.

4. Repeat Item 3 until small positive and small negative values of y_t are obtained. Then by interpolation the correct value of y_m can be calculated.

The above calculations can be repeated for various values of P and t. These data can then be used to determine the load deflection curves with the time t as a parameter (Fig. 4).

For the aluminum alloy tested in this investigation it was found that $n=1$ is a good approximation. As shown in the following, the solution of Eq. (3) is greatly simplified when $n=1$.

SPECIAL THEORY FOR COLUMN DEFLECTIONS WHEN $n=1$

By Eqs. (3) and (b) for $n=1$,

$$\frac{D}{t}\frac{d^2y_c}{dx^2}=-P\left(e+a\sin\frac{\pi x}{L}+y_c\right), \qquad (c)$$

or

$$\frac{d^2y_c}{dx^2}+k^2y_c=-k^2a\sin\frac{\pi x}{L}-k^2e, \qquad (11)$$

where $k^2=Pt/D$.

It can be shown that the solution of this differential equation is

$$y_c=C_1\sin kx+C_2\cos kx+\frac{a\sin\pi x/L}{\dfrac{\pi^2}{k^2L^2}-1}-e. \qquad (12)$$

Boundary conditions can be used to determine the constants C_1 and C_2. Since $y_c=0$ when $x=0$, from Eq. (12),

$$C_2=e. \qquad (d)$$

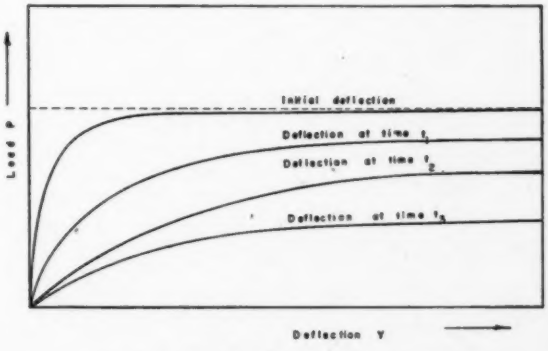


FIG. 4. Load-deflection relations.

[3] J. P. Den Hartog, *Mechanical Vibrations* (McGraw-Hill Book Company, Inc., New York, 1940), p. 225. See also Dana Young, "Inelastic buckling of variable section columns," paper presented at the Annual Meeting of the A.S.M.E., Dec. 1944.

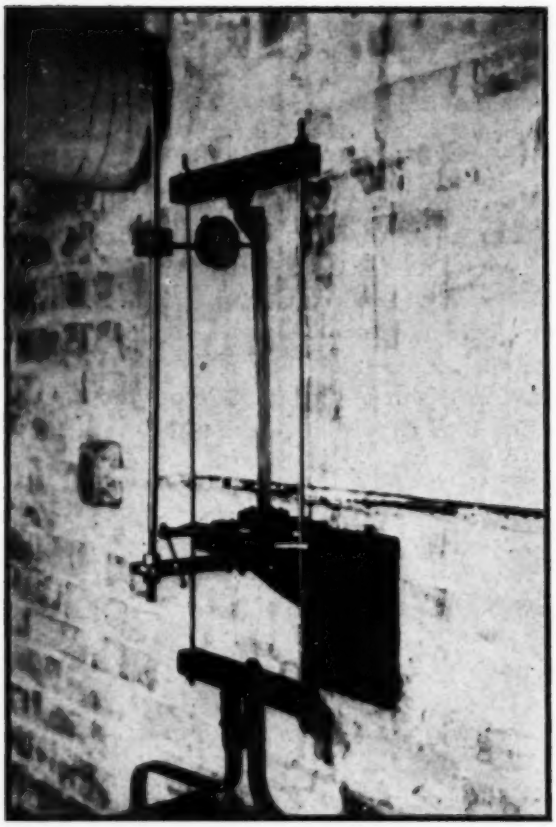


FIG. 5. Creep buckling apparatus.

The deflection y_e also equals zero when $x=L$, or from Eq. (12),

$$C_1=\frac{e(1-\cos kL)}{\sin kL}. \qquad \text{(e)}$$

Substituting the values of C_1 and C_2 from Eqs. (d) and (e) in Eq. (12) the creep deflection is

$$y_c=\frac{\sin kx}{\sin kL}(1-\cos kL)e+e\cos kx-e+\frac{a\sin \pi x/L}{\frac{\pi^2}{k^2L^2}-1}. \qquad (13)$$

By inspection the maximum creep deflection is for a value of $x=L/2$, or

$$y_{\max}=e\left(\sec\frac{PtL^2}{4D}-1\right)+\frac{aPtL^2/D}{\pi^2-\frac{PtL^2}{D}}. \qquad (14)$$

If it is assumed that the initial deflection y_e is entirely the elastic deflection at the center of the column, then it can be shown by the usual elastic theory that,

$$y_e=e\sec[(PL^2/4EI)]^{\frac{1}{2}}-e. \qquad (15)$$

The total deflection at the center (y_T) is equal to the creep deflection ($y_{\max}$), as given by Eq. (14), plus the elastic deflection, as given by Eq. (15). That is, by Eqs. (14) and (15),

$$y_T=e\left[\sec\frac{PtL^2}{4D}+\sec[(PL^2/4EI)]^{\frac{1}{2}}-2\right]+\frac{aPtL^2/D}{\pi^2-\frac{PtL^2}{D}}. \qquad (16)$$

The maximum moment produced in the column at a time t is,

$$M_{\max}=P(y_T+e). \qquad (17)$$

The maximum bending stress with $n=1$ is the same as for the elastic case or equal to $M_{\max}c/I$, where $I=$the moment of inertia of the cross section and $c=$the distance to the outer fiber. In addition to the bending stress, there is an axial compressive stress equal to the load divided by the area or P/A. Then the maximum stress is

$$S_{\max}=\frac{P}{A}+\frac{M_{\max}c}{I}. \qquad (18)$$

Using values of $M_{\max}$ and y_T from Eqs. (16) and (17) in Eq. (18), the maximum stress is

$$S_{\max}=\frac{P}{A}+\left(\frac{Pc}{I}\right)\left[e\sec(PtL^2/4D)+e\sec[(PL^2/4EI)]^{\frac{1}{2}}-e+\frac{aPtL^2/D}{(\pi^2-PtL^2/D)}\right]. \qquad (19)$$

TABLE I. Physical properties of 3S-H aluminum alloy.

Yield strength (0.2 percent set) (pounds per square inch)	25,000
Ultimate strength (pounds per square inch)	29,000
Elongation (percent in 2 inches) (one-half inch round specimen)	10
Hardness (Brinell—500 Kg)	55
Modulus of elasticity (pounds per square inch)	10,300,000
Modulus of rigidity (pounds per square inch)	3,800,000

Equation (19) corresponds to the secant formula obtained in the elastic theory and it can be used to determine maximum stresses in columns accompanied by creep.

EXPERIMENTS ON ECCENTRICALLY LOADED COLUMNS

The columns tested were supported and loaded as shown in Fig. 5. The load was applied to the column specimens by a hanger supplied with a universal joint to prevent bending, and provision was made for rigidly fixing the lower end of the column. The lateral creep deflections were measured by a Federal dial reading to 1/10,000 inch. Precautions were taken for placing the column initially vertical and the dial pointer horizontal. The material tested was an aluminum alloy designated as 3S−H by the Aluminum Company of America and the properties of this material, as supplied by Mr. R. L. Templin, Chief of Tests of this Company, are given in Table I.

Tension and pure bending creep tests also were made on this material at room temperatures.[1] The tension creep tests showed that the creep rate-stress relation was a straight line on a log-log plot. That is, the variation between the stress S and constant creep rate C is:

$$C = AS^n, \tag{20}$$

where A and n are experimental constants.

The values of A and n as determined by the tension tests were reported[1] as $A = 6.51 \times 10^{-12}$ and $n = 1.18$. A reconsideration of this data shows that a value of $n = 1$ and $A = 2.5 \times 10^{-11}$ is also a good approximation. The value of one for the constant n is selected since this simplifies the theoretical interpretation of the test results. The pure bending creep tests showed that Eq. (1) gives a reasonably good approximation for this material.

TABLE II. Dimensions and loads for columns.

Col. No.	Length ($L/2$) in.	Width (b) in.	Thickness (h) in.	Slenderness ratio (L/r)	Eccentricity (e) in.	Load (P) (lb.)	Moment (M) in. lb.
1	14.0	0.50	0.49	195	0.44	258	113.5
2	14.0	0.50	0.50	195	0.79	143	113.0
3	14.0	0.50	0.49	195	0.25	308	77.0
4	14.0	0.50	0.48	195	1.00	113	113.0
5	14.0	0.50	0.49	195	0.64	177	113.3
6	14.0	0.50	0.40	195	0.	433	0.
7	10.2	0.50	0.36	213	0.34	158	53.8
8	10.2	0.50	0.36	213	0.38	158	60.0

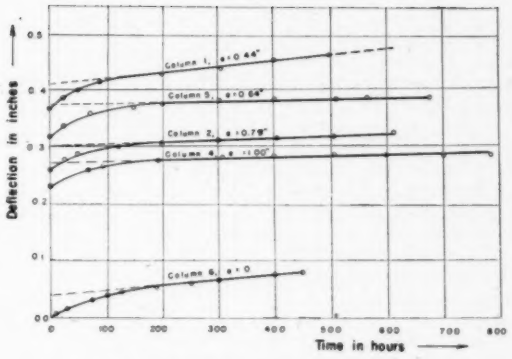


FIG. 6. Deflection-time relations.

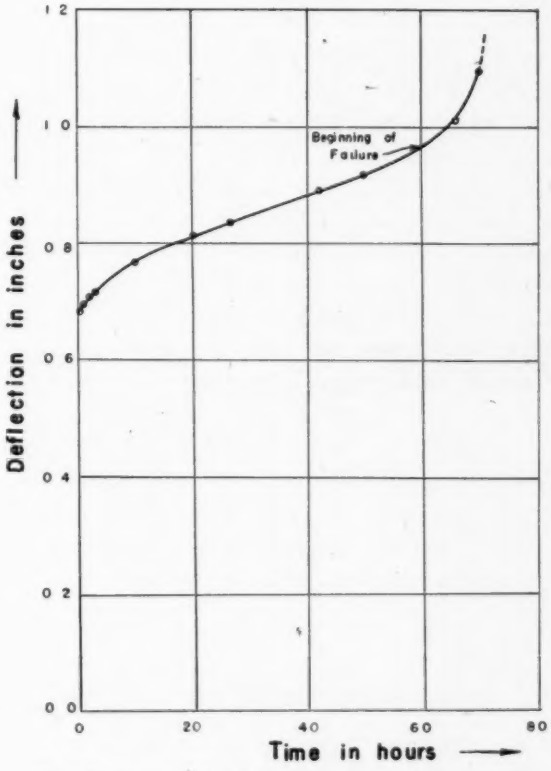


FIG. 7. Creep deflection-time curve for column 3.

The column specimens used were about one-half inch square, and most of them were about 15 inches in length. To prevent local yielding the bases of the columns were increased in size. Various values of the eccentricity e and load P were used for the eight columns tested, as is shown in Table II. Table II also lists the average dimensions and slenderness ratios for the mem-

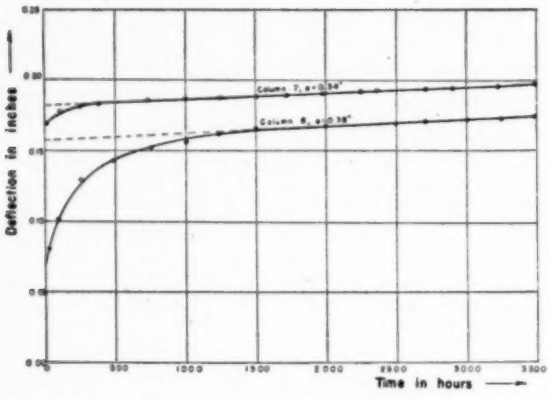


FIG. 8. Deflection-time relations for columns 7 and 8.

bers used. For each experiment, lateral deflections were measured at a point near the upper end of the column, as shown in Fig. 5. An approximate value of the initial elastic and creep deflections was noted for each member. Creep deflection readings were recorded at varying intervals of time and deflection-time plots are shown for the various specimens in Figs. 6–8. Each test covered a time period of at least 500 hours and two specimens covered a time of 3000 hours. After the first 24 hours the creep rate became almost constant so that readings were taken only once a day. The creep rate was constant for all columns except No. 3 which failed by buckling at the end of 70 hours (Fig. 7). For a number of the specimens tested the applied moment was maintained constant by decreasing the load and increasing the eccentricity.

INTERPRETATION OF TEST RESULTS

A comparison between the theoretical and experimental deflection values can be made by determining both the creep deflections and creep rates using each method. The creep deflection given by Eq. (14) also represents the maximum theoretical creep deflection for the column of Fig. 5. For example, to determine the theoretical creep deflection at 500 hours in the case of column 2, it is necessary to substitute in Eq. (14) the values of $e=0.79$ inches, $P=143$ lb., $t=500$ hr., $L=28$ in., $D=I/A=2.08\times10^8$, and $a=0.364$ in. Then the theoretical creep deflection $y_{max}=0.042$ in. The values of the theoretical deflections for other columns are given in Table III.

The actual deflections were not measured at the ends of the columns so that it is necessary to correct these deflections in order to compare them with the theoretical values. To do this the creep deflection curve is assumed to be a sine curve. The corrected values of the measured creep deflections are listed in Table III for comparison with the theoretical values. The values of the creep rates, as obtained by theory and experiment, are also given in Table III for purposes of comparison.

CONCLUSION

In view of the preliminary nature of these tests and the approximations made in the theory, the agreement between the theoretical and experimental creep values as given in Table III is quite satisfactory. There is need, however, for further research on this problem before a definite conclusion can be reached.

ACKNOWLEDGMENTS

The author wishes to express his thanks to R. L. Templin, Chief of Tests of the Aluminum Company of America for the aluminum used in these tests and for control test data supplied. The experimental data of this paper was taken from a Master's Thesis by R. C. DeHart performed under the direction of the writer.

APPENDIX 1

An alternate method for the solution of Eq. (1) is given in the following. The value of y_e is first assumed to be $y_e=ix^2$. The constant i can be determined by noting that $y_e=a$ when $x=L$ (Fig. 9) or $i=a/L^2$. That is,

$$y_e=(a/L^2)x^2, \qquad (f)$$

TABLE III. Comparison of theoretical and experimental values.

Column number	Moment (in. lb.)	Eccentricity (inches)	Creep deflec. (in.) Exp.	Creep deflec. (in.) Theor.	Creep rates (in./in./hr. $\times10^4$) Exp.	Creep rates (in./in./hr. $\times10^4$) Theor.
1	113.5	0.44	0.060	0.057	1.20	1.14
2	113.0	0.79	0.040	0.042	0.80	0.84
4	113.0	1.00	0.030	0.038	0.060	0.076
5	113.3	0.34	0.031	0.043	0.062	0.086
7	53.8	0.34	0.018	0.029	0.060	0.097
8	60.0	0.38	0.018	0.037	0.060	0.123

Note: For columns 1 to 5 creep deflections are for $t=500$ hours. For columns 7 and 8 creep deflections are for $t=3000$ hours.

where a = the initial deflection at B and L = the length of the column. For the column in Fig. 9 the moment is

$$M = Py' = P(e + y_{max} + a - y_c - y_e). \qquad (g)$$

Placing the value of y_e from Eq. (f) and the value of M from Eq. (g) in Eq. (1),

$$d^2y_c/dx^2 = (k_1y_c + k_2x^2 + k_3)^n, \qquad (21)$$

where

$$k_1 = -P(t/D)^{1/n}, \quad k_2 = \frac{-aP}{L^2}(t/D)^{1/n},$$

and

$$k_3 = (e + y_{max} + a)P(t/D)^{1/n}.$$

Placing

$$u = k_1y_c + k_2x^2 + k_3 \qquad (h)$$

in Eq. (18),

$$d^2u/dx^2 = k_1u^n + 2k_2. \qquad (i)$$

Integrating Eq. (i) and separating the variables,

$$\int \frac{du}{(u^{n+1} + k_4u + k_5)^{\frac{1}{2}}} = k_6 \int dx, \qquad (22)$$

where

$$k_4 = (2k_2/k_1)(n+1),$$
$$k_5 = -(k_3)^{n+1} - (n+1)(2k_2k_3/k_1),$$
$$k_6 = [(n+1)/(2k_1)]^{\frac{1}{2}}.$$

Equation (22) can now be solved for particular values of n by using numerical integration methods. For a value of $n = 1$, Eq. (22) can be integrated directly, or

$$u^2 + k_4u + k_5 = (e^{k_6x + k_7} - x - k_4/2)^2, \qquad (23)$$

where k_7 is a constant of integration.

Replacing u by its value in terms of x and y as given by Eq. (h), the equation for the deflection y_c can be obtained.

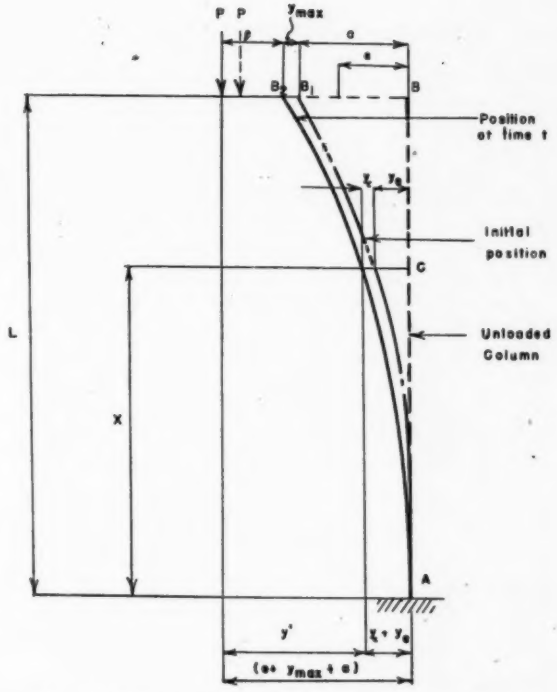


FIG. 9. Deflection of column fixed at one end and free at the other.

Effects of Magnetic Field on Oscillations and Noise in Hot-Cathode Arcs*

J. D. Cobine and C. J. Gallagher**
Radio Research Laboratory, Harvard University, Cambridge, Massachusetts
(Received April 11, 1946)

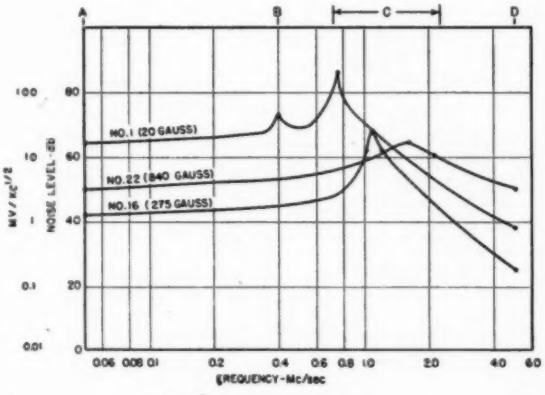
The voltage disturbances in hot-cathode arcs are markedly affected by a magnetic field transverse to the normal flow of current. The normal plasma oscillations can be suppressed and the noise level above 1 Mc/sec. greatly increased. The oscillations appear to be transmitted to the electrodes by electrons whose velocity is modulated by the plasma ion oscillations as they travel directly from cathode to anode. The effect of the magnetic field is to remove these electrons from the current reaching the anode. A critical value of magnetic field will suppress the oscillations and reduce the noise to a minimum. As the magnetic field is further increased the high frequency noise level increases. The high frequency noise level also increases if the pressure decreases, or if a lighter gas is used. The noise level is independent of the plasma volume. The effect of the magnetic field is to build up a region of high ion density in a narrow region near the electron source. Plasma oscillations with continuous spectrum are generated in this region of high density and high density gradient.

INTRODUCTION

OCCASIONALLY mention is made of the fact that gas discharge tubes are "noisy," where "noise" is taken to mean any objectionable fluctuating voltage, such as random noise or oscillations.

Many investigators have reported the existence of oscillations in the voltage appearing between the electrodes of gaseous discharge.[1] The oscillations so observed appear to be of three types, *viz.*, oscillations of electrons and of positive ions in the plasma, and relaxation oscillations. The former appear to be characteristic of the discharge (current, pressure, geometry, nature of gas), while the latter are dependent on both the discharge and the external circuit conditions. Observations and analysis of the plasma oscillations have been made by Tonks and Langmuir.[2] These workers showed theoretically that the frequency of the plasma oscillations increases with increasing positive-ion density. The frequencies of these oscillations have been studied in narrow ranges from 1000 kc/sec to 1000 Mc/sec.

No systematic study of the random noise developed by gaseous discharges has been reported in the literature although gas tubes have been used as random noise generators in acoustical testing,[3] and Tonks and Langmuir[2] mention the existence of a continuum in the voltage output of their discharge tube. Recently the authors have reported a study of random noise and oscillations in hot-cathode arcs.[4] In this study it was shown that the random noise could be explained as the result of random fluctuations of space charge at the electrodes.



Anode current = 30 ma
Load resistance = 2500 ohms

FIG. 1. Selected spectra of hot-cathode gas triode (RCA 884) with transverse magnetic field.

* This work was done in whole or in part under Contract No. OEMsr-411 between the President and Fellows of Harvard College and the Office of Scientific Research and Development.

** Both authors now at Research Laboratory, General Electric Company, Schenectady, New York.

[1] E. V. Appleton and A. G. D. West, Phil. Mag. **45**, 879 (1923). C. Eckart and K. T. Compton, Phys. Rev. **24**, 97 (1924). F. H. Newman, Phil. Mag. **47**, 839 (1924). K. T. Compton and C. Eckart, Phys. Rev. **25**, 139 (1925). R. E. Clay, Phil. Mag. **50**, 985 (1925). L. A. Pardue and J. S. Webb, Phys. Rev. **32**, 946 (1928). E. W. B. Gill, Phil. Mag. **8**, 955 (1929). H. Kniepkamp, Zeits. f. Tech. Phys. **17**, 397 (1936). W. Funk and R. Seeliger, Zeits. f. Physik **113**, 203 (1939).

[2] I. Langmuir, Proc. Nat. Acad. Sci. **14**, 627 (1928); L. Tonks and I. Langmuir, Phys. Rev. **33**, 195 (1929).

[3] E. Peterson, Bell Lab. Record **18**, 81 (1939). H. Dudley, R. R. Riesz, and S. S. A. Watkins, J. Frank. Inst. **227**, 739 (1939).

[4] J. D. Cobine and C. J. Gallagher, Phys. Rev. **70**, 113 (1946). J. Frank. Inst. **243** (Jan. 1947).

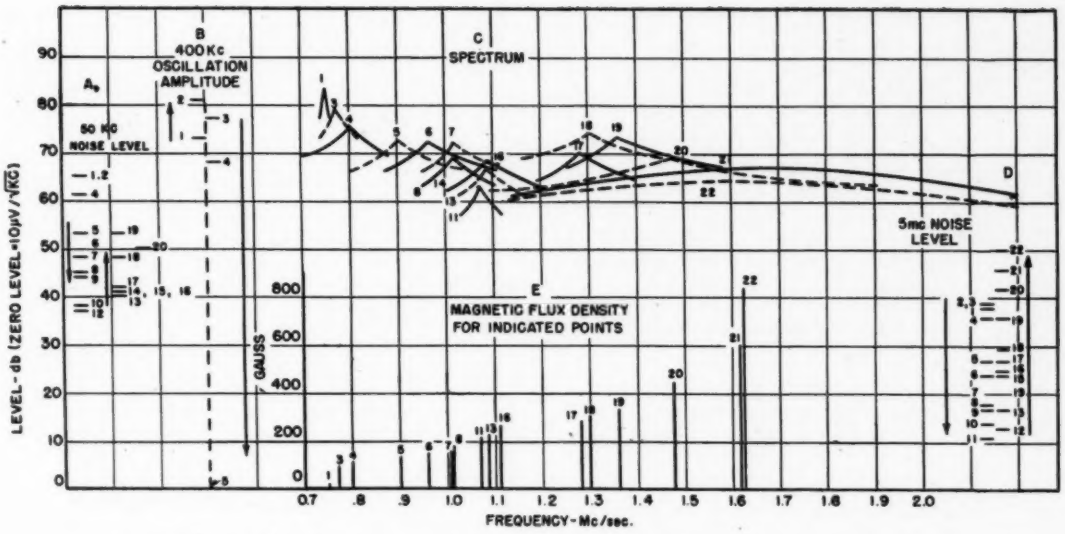


FIG. 2. Effect of magnetic field on noise and oscillations of hot-cathode gas triode (RCA 884).

The present paper presents a further study of noise and oscillations in gas discharges when placed in a magnetic field. A transverse magnetic field is shown to produce two new effects, namely, the suppression of oscillations and the complete alteration of the noise spectrum.

EXPERIMENTAL TECHNIQUE

The voltage fluctuations can be studied effectively by means of spectrum analyzers which measure amplitude and frequency characteristics of the oscillations and random noise. The frequency range from 25 c.p.s. to 9 Mc/sec. could be continuously studied with three heterodyne-type analyzers having band widths sufficiently small that the continuous noise spectrum remained essentially uniform over the region under observation.[5] Thus, the analyzers measured the r.m.s. voltage over a small band of frequencies. Since the r.m.s. noise power in a uniform spectrum is proportional to the band width, all measurements could be referred to a common band width, arbitrarily chosen as 1 kc/sec. The measurements of the spectra under various conditions are presented in db referred to a zero db level of $10\mu v/(kc/sec.)^{\frac{1}{2}}$. Precautions were taken to insure that the input impedance of each analyzer was sufficiently high to avoid "loading" the gas tube output circuit and thus distorting the spectrum being measured. The discharge tube being studied was connected in series with a non-inductive resistance and a regulated power supply. The fluctuation voltage was measured directly between the tube electrodes. The data presented were obtained from both commercial tubes and tubes of experimental design.

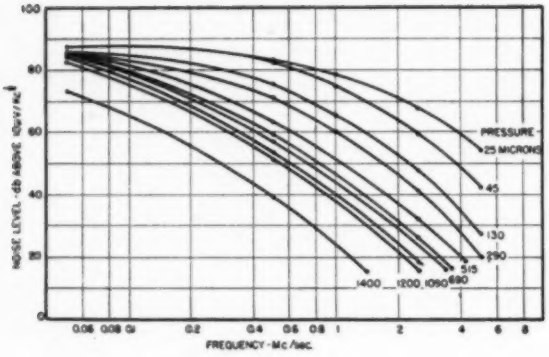
[5] G. P. McCouch and P. S. Jastram, "Video Spectrum Analyzers," Radio Research Lab. Rep., No. 411–96. P. S. Jastram, "Low-Frequency Spectrum Analyzer," Radio Research Lab. Rep., No. 411–154. J. D. Cobine and J. R. Curry, Rev. Sci. Inst. **17**, 190 (1946).

EXPERIMENTAL RESULTS

Oscillations

Much information about the oscillations was conveniently obtained by studying characteristics of type 884 tubes (hot-cathode, argon-filled gas triode). As has been shown previously,[4] the voltage between anode and cathode contains two oscillations superimposed on a background of random noise. These oscillations are not harmonically related.

When a magnetic field is applied transverse to the normal direction of current flow the effect on each oscillation is quite different as shown in Figs. 1 and 2. The effect of the magnetic



Anode current = 10 ma
Magnetic flux density = 860 gauss

FIG. 3. Effect of pressure on spectrum of argon tube.

field on the lower frequency oscillations is indicated in Section B of these figures, where it is seen that as the magnetic field increases, the amplitude first increases and then decreases without any change in frequency. With sufficiently high field (approx. 150 gauss), this oscillation disappears completely. Oscillation C undergoes a more complicated series of changes. The peak shifts toward higher frequencies, with marked broadening. Figure 2, Section C, shows the amplitude of the peak and its spectrum measured between frequencies whose amplitudes are 6 db below those of the peak.

For the strongest magnetic fields, the peak becomes so broad that a discrete oscillation cannot be detected. The numbers on the figure indicate the spectrum levels measured at the corresponding magnetic field in Section E. The behavior of these oscillations is typical of all the hot-cathode arcs studied.

The disappearance of these oscillations was studied in an experimental tube with an oxide-coated cathode and a small plane anode three inches from it. The cathode was partly surrounded by a metal shield to define an electron beam, the limits of which could be plainly seen at a few hundred microns pressure. Under these conditions the discharge space was filled with plasma and the two characteristic oscillations were observed in the voltage spectrum. When a magnetic field was applied so as to deflect the beam of electrons away from the anode, the low frequency oscillation disappeared. The oscillation reappeared at reduced amplitude if the beam was directed against the glass wall of the tube so that electrons coming directly from the cathode were reflected to reach the anode. It is not possible to eliminate this oscillation in any tube in which the electrons could reach the anode directly in spite of the magnetic field. Thus, this oscillation is a positive-ion plasma oscillation whose energy is transmitted to the external circuit by velocity modulation of the electron beam passing directly from cathode to anode. Tonks and Langmuir[5] have shown that the plasma oscillations have zero group velocity and cannot transmit energy directly to the external circuit.

Unlike the plasma oscillation, the second type of oscillation has been observed at all currents, even down into the "Townsend" discharge region where no plasma exists.[4] Similar oscillations have been observed by Ballantine[6] in vacuum tubes containing mercury as a residual gas (pressures less than 1 micron). Since these oscillations appear under conditions in which no plasma is present, as well as under plasma conditions, the most promising explanation appears to be that proposed by Ballantine, *i.e.*, the oscillations are those of positive ions in the potential minimum at the cathode. If these are "cathode" oscillations, the magnetic field has the effect on frequency that might be expected, since the electron density increases near the cathode due to the bending of the electron paths in the magnetic field. The increase in density would cause a deepening of the potential well, with consequent increase in oscillation frequency. The broadening of the peak is more difficult to explain.

Noise

As shown in Fig. 1, the oscillations are superimposed on a continuous background of random noise. The effect of the magnetic field is first to

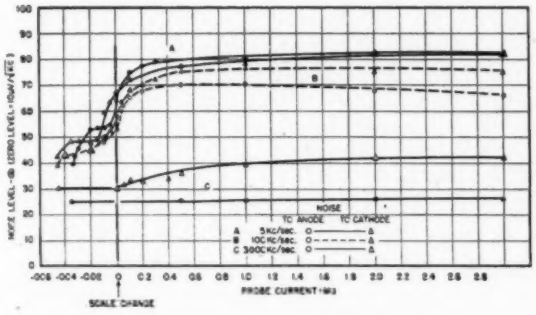
TABLE I

Noise characteristics of hot-cathode arcs in different gases (1000 microns, 10 ma, 720 gauss)

Gas		He	A	Xe	Hg*
Noise level (db above $10\mu v/kc^{\frac{1}{2}}$)	50 kc/sec.	89	73	83	80
	1 Mc/sec.	81	69	63	62
	5 Mc/sec.	48	42	32	33
Peak-to-peak voltage		32–37	16	11.6	—
Arc drop (volts)		48	20	12	
Ionization potential (volts)		24.5	15.69	12.08	10.38
Atomic wt.		4.002	39.94	131.3	200.6

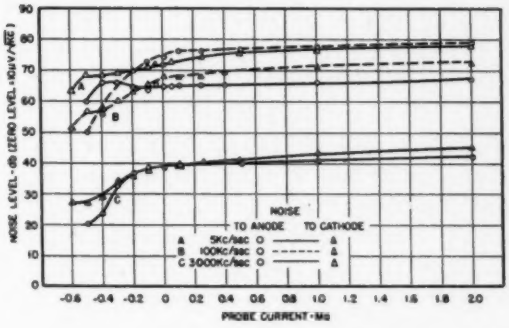
* 125 ma.

[6] Stuart Ballantine, Physics **4**, 294 (1933).



Magnetic flux density = 800 gauss at cathode
Cathode current = 20 ma
Argon at 85 microns
Δ —noise measured between anode probe and cathode
○—noise measured between anode probe and anode

FIG. 4. Effect of probe current on noise at anode probe.



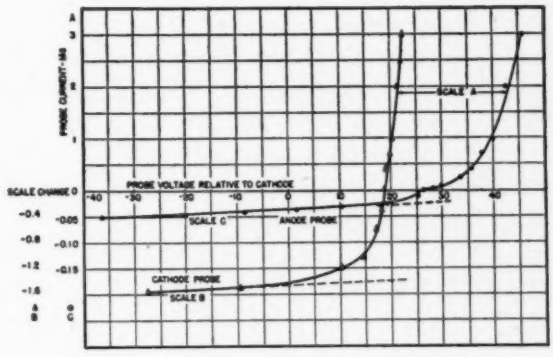
Magnetic flux density = 800 gauss at cathode
Cathode current = 20 ma
Argon at 85 microns
Δ —noise measured between cathode probe and cathode
○—noise measured between cathode probe and anode

FIG. 5. Effect of probe current on noise at cathode probe.

decrease and then increase the noise level over the entire spectrum. In no case where the magnetic field is equally strong at cathode and anode is the low frequency noise increased above the zero field value. On the other hand the high frequency noise can be increased markedly over the zero field value (Sections A and D, Fig. 2). The noise level in the high frequency region increases continuously up to the point where the magnetic field is strong enough to extinguish the discharge. The ultimate noise voltage level at 5 Mc/sec. may be 100 times the zero field level.

The noise level is also affected by gas pressure. With the discharge in a strong magnetic field, the noise increases continuously as the pressure is reduced, to the point where the pressure is too low for the discharge to be maintained in the presence of the field. Figure 3 shows a series of typical spectra obtained with an argon-filled experimental tube with a field of 860 gauss at the cathode. It will be noted that there is no "hump" in these spectra. This characteristic was noted for all tubes in which the electrode structure was such that the field was strong at one electrode only. In this case, the hump which is present at lower fields is obliterated by the high level of the low frequency noise.

The discharge current does not greatly effect the shape of the spectrum, if arc conditions are fully established. Both high and low frequency noise increase with increasing currents by a factor of not more than 3 when the arc current increases by a factor of 100. The noise is also substantially independent of the plasma volume, which was varied by a factor of 2000.



Magnetic flux density = 800 gauss at cathode
Cathode current = 20 ma
Argon at 85 microns

FIG. 6. Probe characteristics of experimental tube (Figs. 4 and 5).

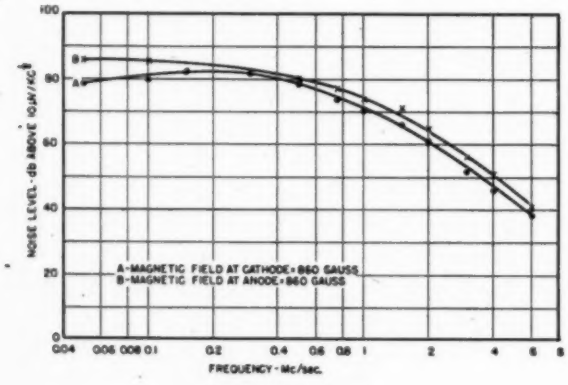
The effect of different gases on the noise was studied. The general shape of the spectra is the same for all of the gases used, including helium, argon, xenon, and mercury. The high frequency noise (above 1 Mc/sec.) decreased as the atomic weight increased. No definite trend was noted in the low frequency noise. The results of these studies are summarized in Table I. The peak-to-peak voltage was measured on a wide-band oscilloscope. The peak-to-peak voltage is of the order of 85 percent of the arc drop. The arc drop itself is roughly proportional to the ionization potential of the gas used.

It was found that the amount of noise voltage measured between either cathode or anode, and an auxiliary probe electrode was highly dependent on the current collected by the probe. For these measurements, the main discharge was

maintained between cathode and anode. The cathode-anode current and noise voltage were unaffected by the variations in the probe current. Two probes were used—one near the cathode and the other near the anode. The magnetic field was applied at the cathode and was perpendicular to the plane of the cathode probe. For each probe the noise was measured between probe and anode and probe and cathode. Figures 4 and 5 show the noise level at three frequencies as a function of probe current. The volt-ampere characteristics of the probes are shown in Fig. 6.

In all cases the noise was very low when the probe was made sufficiently negative that only positive-ion current was drawn. In general, the noise increases only when the probe begins to draw electron current. The noise at the cathode probe changes very gradually with current. The noise at the anode probe changes very abruptly when appreciable electron current is drawn. With magnetic field, the noise reaches its greatest value when the probe potential is such as to repel all but the fastest electrons. For these probe currents, conditions are such that very little noise is developed in the space charge region adjacent to the probe, since without the magnetic field, it has been shown that the maximum noise at the probe is developed only when its voltage is such as to accelerate electrons.[4]

Since the plasma acts as a "virtual" cathode with respect to the anode, it seemed probable that placing the magnetic field at the anode should produce the same effect on the noise spectrum as a field at the cathode. The experiment was tried, and the results were as expected (Fig. 7). In fact, the noise with the field at the anode was actually of higher level than that produced by a field at the cathode. Placing the magnetic field at any other position in the discharge decreased the over-all noise level, *i.e.*, the magnet had to be placed at a source of electrons to produce the high level noise. It has been observed elsewhere that the light from the discharge fluctuates with a "light" noise spectrum identical in shape with that of the electrical noise spectrum.[7] The light fluctuations were studied with a photo-cell. This effect was also noted in our investigation.



Argon at 85 microns
Anode current = 20 ma

FIG. 7. Noise spectrum of experimental tube for magnetic field at cathode (A) and at anode (B).

With a sufficiently strong magnetic field at the cathode, there is a narrow region of intense illumination adjacent to the cathode. Numerous experiments with fixed and moving probes show that the positive-ion density is abnormally high in the region close to the cathode when the magnetic field is applied. The positive-ion density was measured by the method of Langmuir and Mott-Smith.[8] For example, in one tube the positive-ion density near the cathode changed from 0.6×10^{10} ions/cm^3 to 3.0×10^{10} ions/cm^3 upon application of the magnetic field.

DISCUSSION

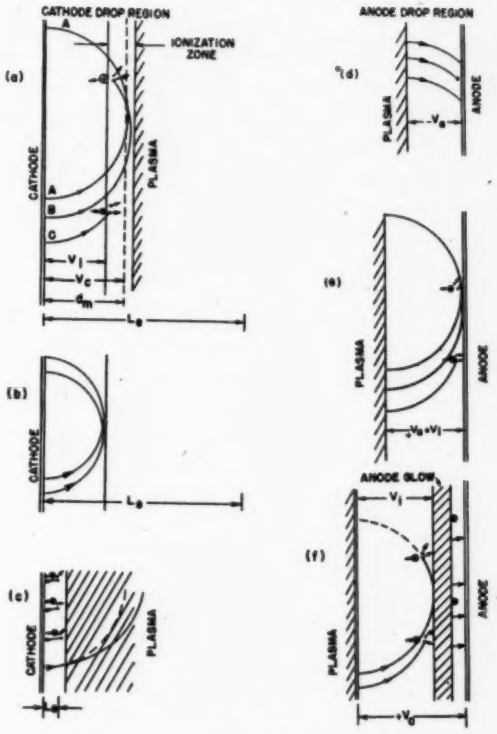
It is apparent that the generation of noise in the presence of the magnetic field is closely related to the space charge conditions at the cathode, or in particular, to the increased positive-ion density mentioned above. That the magnetic field must increase the positive-ion density in a narrow region near the cathode can be seen by considering Fig. 8. This figure presents a picture of the processes at a plane cathode and anode where each is in a strong magnetic field. The trajectory of an electron which does not collide with any atoms (A of Fig. 8a) will return to the cathode. The path will be determined by the electric and magnetic fields. In a uniform magnetic field, with plane electrodes and with no space charge or initial velocity, the path is cycloidal. If the electric field is distorted by

[7] Observed by Dr. R. W. Engstrom, Radio Corporation of America, Lancaster, Pennsylvania.

[8] I. Langmuir and H. Mott-Smith, Gen. Elec. Rev. **27**, 449 (1924).

space charge, a direct calculation of the path is not possible unless a relation between electric field and distance is known. Thomson has shown that the paths in the dark space of a glow discharge in a transverse magnetic field are similar in form to cycloids.[9] It is probable that the paths do not differ greatly in the hot-cathode arc. If the pressure is such that the electrons make most of their collisions when near the point of maximum distance from the cathode, it is apparent that an intensely ionized zone will be found at this distance. If there is a very small spread in the initial velocities of the electrons, this zone will be very narrow. At higher pressures, the electrons will ionize much closer to the cathode, and the magnetic field will have little effect. Also, since the electric field is greater, the projected path will extend farther from the cathode (Fig. 8c). As the pressure is reduced, the point of maximum ionization moves out along the trajectory. At the same time, the maximum in the trajectory moves nearer the cathode because of the smaller electric field. At very low pressures the ionization must be occurring in a very narrow region, since reduction of the pressure makes it more and more probable that a collision will not occur until the electron has passed the maximum in the trajectory. In this portion of the path the electron is losing energy to the field. Thus, the electron must ionize before it reaches the point where its energy is less than the ionization potential (B of Fig. 8a). In the limiting case, immediately preceding extinction of the discharge, the zone of ionization must be very narrow. This picture assumes that most of the ionization results from single collisions. Multiple ionization, although certainly occurring, is not favored under conditions of low pressures. Such cumulative ionization as does occur would broaden the ionization zone.

It is also possible for the magnetic field to establish a narrow zone of ionization at the anode. Possible space charge conditions are shown in Fig. 8 (d, e, f). If the anode drop is negative, electrons coming from the plasma are retarded, and little ionization occurs near the anode (Fig. 8d). If the anode drop is positive



(a) Cathode region, pressure = few hundred microns
A = trajectory of electron making no collisions
B = trajectory of electron making collision as it leaves ionization zone
C = trajectory of electron making collision as it enters ionization zone
(b) Cathode region, same pressure as (a), magnetic flux density = cut off value
(c) Cathode region, high pressure, same magnetic flux density as in (a)
Dotted trajectory same as A of (a)
Solid trajectory at high pressure
(d), (e), (f) Anode region, same pressure and magnetic flux density as in (a)
L_e = electron mean free path
V_i = ionization potential of gas
V_c = cathode drop of potential
d_m = mean distance of ionization zone from cathode
V_a = anode drop of potential

FIG. 8. Details of electron paths with transverse magnetic field.

and slightly greater than the ionization potential, space charge conditions are similar to those at the cathode, since the plasma acts as a source of electrons (Fig. 8e). If the anode is small, a second ionized zone may be formed adjacent to it (Fig. 8f).

Dr. Felix Bloch suggested during the course of this research that the source of the increased noise is in the region of high space-charge density, where the ions in each element of volume can oscillate with a frequency determined by the local ion density, as in the plasma oscillations studied by Tonks and Langmuir. Since the ion density varies greatly over this region, a wide range of frequencies would be present instead

[9] J. J. Thomson, *Conduction of Electricity through Gases* (Cambridge University Press, London, 1933), third edition, Vol. II, p. 333 ff.

of the single characteristic frequency of the plasma. Tonks and Langmuir have shown that the plasma type oscillation can take place if the mean distance between ions is less than the Debye distance. For the ion densities and temperatures observed here, the Debye distance is several orders of magnitude greater than the interionic spacing. An unpublished analysis by Drs. Felix Bloch and Robert Weinstock[10] indicates that the highly ionized region may be as thin as 5×10^{-5} cm and is capable of generating frequencies as high as 250 Mc/sec.

Additional evidence for the effect of the narrow ionization zone in producing the noise may be found in some additional experiments with oxide-coated cathodes. A number of commercial tubes were found deficient in high frequency noise unless the cathodes were properly treated to insure good emission. As might be expected, the arc drop was higher than for a normal tube at the same current. The higher arc drop is known to result in sputtering of the cathode materials which have much lower ionization potentials than argon. (Ionization potential of barium = 5.19 volts.) These sputtered atoms can be ionized over a broad region beginning close to the cathode. Thus, if there is an appreciable density of sputtered atoms, there will be a marked broadening of the ionization zone, with consequent lowering of the ion-density peak. This principle of the variation of high frequency noise level with cathode activation was used in designing an instrument for production testing of commercial tubes.

[10] Radio Research Lab. Rep., No. 411–232.

ACKNOWLEDGMENT

The authors wish to acknowledge the generous assistance of Dr. J. H. Van Vleck and Dr. Felix Bloch for many consultations during the course of this research. Acknowledgment is also made to Dr. Martin Freundlich and Mr. Stanley Ruthberg who built most of the experimental tubes used for these studies.

Numerical Solution of Axisymmetrical Problems, with Applications to Electrostatics and Torsion

GEORGE SHORTLEY, *Mendenhall Laboratory of Physics, Ohio State University, Columbus, Ohio*

ROYAL WELLER, *Naval Ordnance Laboratory, Washington, D. C.*

PAUL DARBY,* *Mendenhall Laboratory of Physics, Ohio State University, Columbus, Ohio*

AND

EDWARD H. GAMBLE,** *Mendenhall Laboratory of Physics, Ohio State University, Columbus, Ohio*

(Received September 19, 1946)

Numerical methods are given for solution of axisymmetrical problems involving the partial differential equation

$$\frac{\partial^2\psi}{\partial z^2}+\frac{\partial^2\psi}{\partial \rho^2}+\frac{K}{\rho}\frac{\partial\psi}{\partial\rho}=0,$$

where ρ is the radial coordinate and z the coordinate parallel to the axis. The various values of K which occur in physical situations are discussed, and common iteration methods for handling these problems are given. For Laplace's equation, $K=1$. For the Stoke's stream function, $K=-1$, but it is pointed out that for numerical work a new function, called the *flow-disturbance function*, having $K=3$, is more tractable. Similarly a new function with $K=5$, the *stress-concentration function*, is much easier to compute than the usual stress function ($K=-3$) for the case of the torsion of a circular shaft of varying diameter. The methods are illustrated by computation of the equipotentials for an electron lens, and by a complete computation of the stresses and strains in a particular grooved circular shaft under torsion.

* Now at Remington Arms Company, Bridgeport, Connecticut.
** Now at Bell Telephone Laboratories, Murray Hill, New Jersey.

1. INTRODUCTION

THIS paper represents an extension of the numerical methods for handling partial differential equations in two dimensions, which were developed in recent work on "Numerical Solution of Laplace's and Poisson's Equations, with Applications to Photoelasticity and Torsion." [1] For the basis of some of the methods used in the present paper, reference should be made to this previous work. On the other hand, the method of extrapolation developed in Section 6 of the present paper is an improvement over that previously given, and can be used in the solution of Laplace's and Poisson's equations in two dimensions as well as in the cases discussed here.

An axisymmetrical problem is essentially two-dimensional, since the functions which arise depend only on the coordinate z, measured parallel to the axis, and the radius ρ, measured normal to the axis, and are independent of the azimuthal angle φ. The differential equations which govern many functions of technical interest are of the form

$$\frac{\partial^2\psi}{\partial z^2}+\frac{\partial^2\psi}{\partial \rho^2}+\frac{K}{\rho}\frac{\partial\psi}{\partial\rho}=0, \qquad (1)$$

which we abbreviate to

$$\Delta'\psi+(K/\rho)\psi_\rho=0. \qquad (1')$$

Here the constant K takes on the following values:

$K=1$, Laplace's equation, applicable to the potentials of axially symmetric electrostatic, magnetostatic, and gravitational fields; and the steady flow of heat, electric current, and ideal fluid. This case is discussed in Section 7, with an application to an electron-lens system.

$K=-1$, Stoke's stream function, which is constant along a stream line in the steady flow of heat, electric current, and ideal fluid. This function, while it satisfies simpler boundary conditions than the potential, has the disadvantage for numerical work of rapid variation with ρ; for a unidirectional stream it varies as ρ^2. Hence it is preferable to introduce a new function, the *flow-disturbance function*, which is the stream function divided by ρ^2. This function is constant in regions free from disturbance and satisfies the above equation with

$K=3$; the use of the flow-disturbance function for numerical work is exactly analogous to the use of the stress-concentration function which is discussed in detail in this paper.

$K=-3$, a function which governs the stresses in a circular shaft of varying diameter. We introduce a new function, the *stress-concentration function*, more suitable for numerical work, which satisfies the equation with

$K=5$; the use of the stress-concentration function is the subject of Sections 8–10 of this paper.

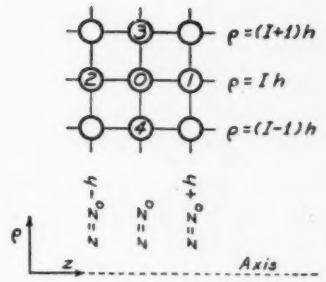


FIG. 1. Notation for an ordinary point of the net.

Because these problems differ only in the value of K in (1), the methods of numerical solution will be similar and can be discussed without specifying the value of K.

It is well to point out at this time that any numerical integration of an elliptic partial differential equation in two or more dimensions is bound to be very tedious and time-consuming if high accuracy is desired simply because of the number of net points which must be used. The more irregular the boundaries, the more points which must be used if accuracy is to be obtained in the vicinity of the irregularities. But it seems important to know that a computation can be carried through numerically when required, and to investigate the most expeditious manner of making the computation. The procedures given below and in reference 1 are believed to contain most of the formal tricks for speeding up the computations which have been discovered to date and to constitute the best available computational scheme. It is hoped that some day high speed electronic computing equipment will be developed with sufficient internal-memory capacity to handle data on a large net; in this case it may be advisable to ignore tricks, use the simplest possible improvement formula at each point, and let the machine work somewhat

[1] G. Shortley and R. Weller, J. App. Phys. 9, 334 (1938); J. App. Mech. 6, A-71 (1939); Weller, Shortley, and Fried, J. App. Phys. 11, 283 (1940). These three papers have been combined in the Ohio State University Engineering Experiment Station Bulletin No. 107 (1940); revised, 1942. In these papers will be found complete references to earlier work.

longer, or to use some non-iterative procedure such as a direct numerical solution of the difference equations.

2. THE ITERATION PROCEDURE

We first replace the continuum of points (z, ρ) by the discrete set of points at the intersections of a net of intervals h:

$$z=0,\ \pm h,\ \pm 2h,\ \pm 3h,\ \cdots,$$
$$\rho=0,\ h,\ 2h,\ \cdots,\ Ih,\ \cdots.$$

The differential equation is then replaced by a difference equation which relates the value of ψ at one point to its values at neighboring points. The solution of the difference equation is then an approximation to the solution of the differential equation—the finer the net, the better the approximation.

For example, consider the five numbered points in the diagram (Fig. 1). Denote the function values at these points by ψ_0, ψ_1, ψ_2, ψ_3, and ψ_4. At the point 0 ($z=z_0$, $\rho=Ih$), Eq. (1′) may be written as

$$\Delta'\psi_0+(K/Ih)\psi_{\rho 0}=0$$

or

$$h^2\Delta'\psi_0+(K/I)h\psi_{\rho 0}=0. \qquad (2)$$

Now on any quadric surface passing through the five function values $\psi_0, \cdots, \psi_4$, the following relations are satisfied:

$$h^2\Delta'\psi_0=-4\psi_0+\psi_1+\psi_2+\psi_3+\psi_4,$$
$$h\psi_{\rho 0}=\tfrac{1}{2}\psi_3-\tfrac{1}{2}\psi_4.$$

Hence the differential equation (2) is approximated—to the accuracy with which the function ψ is representable by a quadric in the region containing these five points—by the difference equation

$$-4\psi_0+\psi_1+\psi_2+\psi_3+\psi_4+(K/I)(\tfrac{1}{2}\psi_3-\tfrac{1}{2}\psi_4)=0 \qquad (3)$$

or

$$8I\psi_0=2I\psi_1+2I\psi_2+(2I+K)\psi_3+(2I-K)\psi_4. \qquad (4)$$

This difference equation holds at all ordinary points of the net, but needs modifications, which will be considered later, at points near the axis or near a material boundary.

The iteration procedure, originated by Liebmann,[1] for finding the function values which satisfy this difference equation at all net points, proceeds as follows: Start with any arbitrary "trial" function ψ—a guess. Then go over the points of the net, in some definite order, replacing the ψ value at each net point by a new value calculated from (4) in terms of the values at neighboring points. This will give an "improved" function ψ'. In calculating the improved value at a given point, the previously altered values are used whenever they are available. This procedure is then iterated. Successive traverses of the net give successively functions $\psi', \psi'', \psi''', \cdots$ which approach the true solution of the difference equation in a manner to be considered in Section 6.

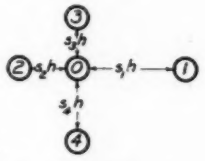
We shall now consider modifications of the difference equation (3), necessary at points near the axis or near a boundary, and modifications leading to more rapid convergence. The essence of the iteration procedure is that at any point the difference equation, solved for ψ_0 as in (4), may be used as an "improvement formula." A proof of convergence, by this process, to the exact solution of the difference equation, and a method of investigating the rate of convergence, are developed in reference 1. This theory assumes that the improvement formula expresses the value at a point as a *true weighted average* of the neighboring values, *with all weights positive*. We note that the coefficient of ψ_0, in (4) and the following improvement formulas, equals the sum of the coefficients of the neighboring values ψ_1, ψ_2, $\cdots$. A little experimentation shows that if negative coefficients are permitted, the procedure may not converge, so we reject any improvement formula which involves negative coefficients and attempt to replace it by an alternative formula in which all the coefficients are positive.

3. IMPROVEMENT FORMULAS

The following formulas are derived by passing a quadric through $\psi_0, \psi_1, \psi_2, \cdots$ which is required to satisfy the differential equation at point 0. The mechanism for accomplishing this is discussed in Appendix 1. Certain of the coefficients in these formulas, as well as in (4), become negative for high K and low I values. Formulas with negative coefficients may be used for interpolation but not for improvement. Modifications

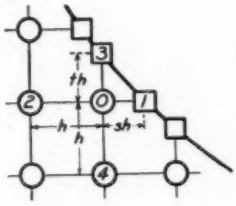
necessary to eliminate these negative coefficients will be considered later for each K value separately. *In each diagram below, the coordinate axes are assumed to run as in Fig. 1, with the symmetry axis horizontal and below the diagram, and with coordinates $z=z_0$, $\rho=Ih$, for the point 0.*

Unequal spacing:



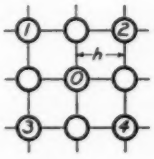
$$\left(\frac{2I}{s_1s_2}+\frac{2I+K(s_4-s_3)}{s_3s_4}\right)\psi_0=\frac{2I}{s_1(s_1+s_2)}\psi_1+\frac{2I}{s_2(s_1+s_2)}\psi_2 +\frac{2I+Ks_4}{s_3(s_3+s_4)}\psi_3+\frac{2I-Ks_3}{s_4(s_3+s_4)}\psi_4. \quad (5)$$

Point near boundary:



$$\left(\frac{2I}{s}+\frac{2I+K(1-t)}{s}\right)\psi_0=\frac{2I}{s(1+s)}\psi_1+\frac{2I}{1+s}\psi_2 +\frac{2I+K}{t(1+t)}\psi_3+\frac{2I-Kt}{1+t}\psi_4. \quad (6)$$

Diagonal formula:

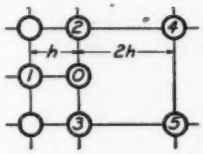


$$8I\psi_0=(2I+K)\psi_1+(2I+K)\psi_2 +(2I-K)\psi_3+(2I-K)\psi_4. \quad (7)$$

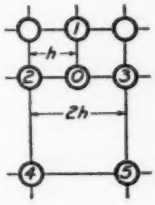
Diagonal formula, unequal spacing:

$$\left(\frac{2I+K(s_4-s_1)}{s_1s_4}+\frac{2I+K(s_3-s_2)}{s_3s_2}\right)\psi_0=\frac{2I+Ks_4}{s_1(s_1+s_4)}\psi_1 +\frac{2I+Ks_3}{s_2(s_2+s_3)}\psi_2+\frac{2I-Ks_2}{s_3(s_2+s_3)}\psi_3+\frac{2I-Ks_1}{s_4(s_1+s_4)}\psi_4. \quad (8)$$

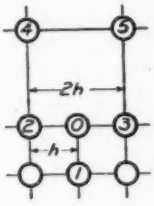
Change of net spacing:



$$16I\psi_0=4I\psi_1+(5I+3K)\psi_2+(5I-3K)\psi_3+I\psi_4+I\psi_5; \quad (9)$$

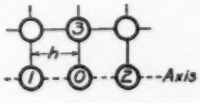


$$(32I+8K)\psi_0=(8I+8K)\psi_1+(10I+K)\psi_2 +(10I+K)\psi_3+(2I-K)\psi_4+(2I-K)\psi_5; \quad (10)$$

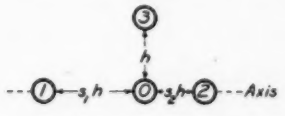


$$(32I-8K)\psi_0=(8I-8K)\psi_1+(10I-K)\psi_2 +(10I-K)\psi_3+(2I+K)\psi_4+(2I+K)\psi_5. \quad (11)$$

Axis points (obtained from a quadric, even in ρ, which satisfies the differential equation everywhere):

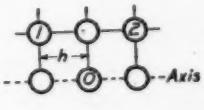


$$(2K+4)\psi_0=\psi_1+\psi_2+(2K+2)\psi_3; \quad (12)$$



$$\left(1+K+\frac{1}{s_1s_2}\right)\psi_0=\frac{1}{s_1(s_1+s_2)}\psi_1+\frac{1}{s_2(s_1+s_2)}\psi_2 +(1+K)\psi_3. \quad (13)$$

Axis point, diagonal (obtained from the linear solution $\psi=\psi_0+Az$; not to be used except for interpolation):



$$\psi_0=\tfrac{1}{2}(\psi_1+\psi_2). \quad (14)$$

4. NINE-BLOCK FORMULAS

As explained in reference 1, the convergence of the iteration procedure is much more rapid if we improve, not one point at a time, but a block of points all at once. A block of nine points is the maximum that can be handled conveniently. In the case of Laplace's equation, the use of nine-blocks was found to cut the labor by a factor of 3.5, and the saving of time is of the same order of magnitude for these axially symmetric problems.

The idea is to find improved values for all the nine circled points of Fig. 2 in terms of the

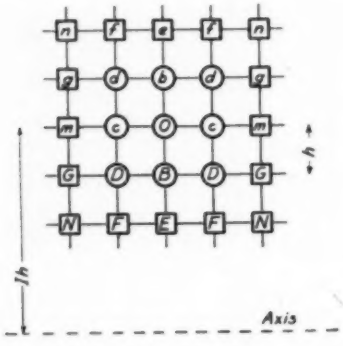


FIG. 2. The nine-block.

values at the surrounding points, marked by squares. We start by writing the difference equation (4) for each of the nine points in terms of its neighbors. Elimination of the ψ values at points b, B, c, d, D then leaves a formula for ψ_0 in terms of only the outside points e, E, f, F, g, G, m. This formula gives an improved value for ψ_0 in which the difference equation is satisfied simultaneously at all nine points of the block. The coefficients in this improvement formula are fifth-order polynomials in I and K (see Appendix 2), so we give in Tables I and II the numerical values needed for $K=1$ and $K=5$. These tables give the coefficients by which each of the 12 boundary points are to be multiplied to get the improved center-point value. The coefficients given for points F, G, m, g, and f are each to be used twice.

After the center value ψ_0 has been determined in this way, the four corner values d, D, are determined from it and the outside values at e, E, m, n, N by using the diagonal formula (7); finally the side points b, B, c are obtained from the ordinary improvement formula (4). In this way improved values for all nine points of the block are filled in without using any of the old values at these points.

TABLE I. Coefficients for improvement of nine-block center point ($K=1$).

	Coefficients of boundary points in Fig. 2						
I	E	F	G	m	g	f	e
0	—	—	—	0.042553	0.106383	0.159574	0.382979
2	0.048393	0.024451	0.048902	0.131764	0.081503	0.095087	0.188193
3	.079053	.039681	.052908	.127476	.074072	.083331	.166010
4	.091719	.045955	.055147	.126312	.070903	.077993	.155661
5	.098824	.049476	.056544	.125818	.069110	.074869	.149543
6	.103402	.051747	.057497	.125560	.067951	.072805	.145479
7	.106607	.053338	.058187	.125408	.067139	.071335	.142578
8	.108979	.054517	.058710	.125311	.066538	.070235	.140400
9	.110807	.055425	.059120	.125244	.066075	.069379	.138704
10	.112259	.056147	.059450	.125198	.065708	.068695	.137346
11	.113441	.056735	.059721	.125163	.065409	.068134	.136233
12	.114422	.057223	.059948	.125137	.065161	.067667	.135305
13	.115249	.057635	.060141	.125116	.064952	.067272	.134519
14	.115955	.057987	.060307	.125100	.064774	.066933	.133844
15	.116567	.058291	.060450	.125087	.064619	.066639	.133259
16	.117100	.058557	.060577	.125077	.064485	.066381	.132746
17	.117570	.058792	.060688	.125068	.064366	.066154	.132294
18	.117988	.059000	.060787	.125060	.064261	.065952	.131891
19	.118361	.059185	.060876	.125054	.064167	.065771	.131531
20	.118696	.059352	.060957	.125049	.064083	.065608	.131206
21	.118999	.059504	.061029	.125044	.064006	.065461	.130913
22	.119274	.059641	.061095	.125040	.063937	.065327	.130645
23	.119525	.059766	.061156	.125037	.063874	.065205	.130401
24	.119755	.059881	.061211	.125034	.063816	.065092	.130177
25	.119966	.059986	.061262	.125031	.063763	.064989	.129971
26	.120161	.060083	.061309	.125029	.063714	.064894	.129781
27	.120341	.060173	.061353	.125027	.068668	.064805	.129605
28	.120509	.060257	.061394	.125025	.063626	.064723	.129441
29	.120665	.060335	.061432	.125023	.063587	.064647	.129289
30	.120810	.060407	.061467	.125022	.063551	.064576	.129147

TABLE II. Coefficients for improvement of nine-block center point ($K=5$).

	Coefficients of boundary points in Fig. 2						
I	E	F	G	m	g	f	e
0	—	—	—	0.043478	0	0.130435	0.652174
4	0.007149	0.003466	0.020797	0.107452	0.090121	0.135182	0.278813
5	.022148	.010862	.028965	.113597	.086895	.123101	.251011
6	.035056	.017296	.034593	.117015	.084011	.114015	.231084
7	.045545	.022552	.038661	.119104	.081618	.107123	.216339
8	.054042	.026822	.041722	.120471	.079652	.101777	.205069
9	.060996	.030320	.044102	.121414	.078027	.097534	.196210
10	.066762	.033223	.046002	.122090	.076669	.094094	.189080
11	.071605	.035663	.047551	.122592	.075522	.091256	.183225
12	.075724	.037738	.048838	.122975	.074542	.088877	.178336
13	.079265	.039522	.049923	.123273	.073696	.086856	.174196
14	.082339	.041071	.050850	.123510	.072958	.085118	.170645
15	.085032	.042427	.051651	.123702	.072311	.083609	.167569
16	.087409	.043624	.052349	.123859	.071738	.082287	.164877
17	.089522	.044688	.052964	.123989	.071227	.081120	.162504
18	.091412	.045640	.053509	.124098	.070769	.080081	.160395
19	.093112	.046496	.053995	.124190	.070357	.079152	.158509
20	.094650	.047270	.054432	.124269	.069983	.078315	.156814
21	.096047	.047973	.054826	.124337	.069644	.077558	.155280
22	.097322	.043614	.055183	.124396	.069333	.076869	.153888
23	.098490	.049201	.055509	.124447	.069048	.076241	.152617
24	.099564	.049741	.055808	.124492	.068786	.075665	.151453
25	.100555	.050240	.056081	.124532	.068544	.075135	.150382
26	.101471	.050700	.056334	.124567	.068320	.074646	.149395
27	.102322	.051128	.056567	.124599	.068112	.074193	.148481
28	.103113	.051526	.056784	.124627	.067918	.073773	.147634
29	.103852	.051897	.056985	.124652	.067737	.073381	.146845
30	.104542	.052244	.057172	.124675	.067567	.073016	.146109

5. USE OF THE DIFFERENCE FUNCTION

A great saving in labor is achieved by carrying out operations, not on the function ψ itself, but on the "difference function," which represents the change in ψ with one traverse. We suppose that the net has been completely laid out, the improvement formula for each point and a definite order of traversing the points and blocks decided on, and the initial trial function chosen. Suppose that the net were traversed several times, to obtain successively the improved functions ψ', ψ'', $\cdots$. We call the change in function value at each point, $\delta=\psi'-\psi$, the *difference function*. Let δ' be the next change or difference: $\delta'=\psi''-\psi'$. Then if, after one traverse, we compute δ at each point by subtraction, we can compute what the next change δ' will be by applying our improvement formulas directly to δ. For, application of the improvement process to ψ gives ψ', application to ψ' gives ψ'', application to the difference $\delta=\psi'-\psi$ will give the difference $\delta'=\psi''-\psi'$.

Hence after *one* traverse, we can compute all successive changes which ψ will undergo by operating on the difference function δ in place of operating on ψ itself. There are a number of factors which make it much simpler to work with the difference function than with ψ. First, the difference function usually has fewer significant figures than has ψ. Second, the boundary values of the difference function are all zero; this gives many zero terms in the improvement formulas. Finally, we know the value to which the difference function will converge after a large number of improvements, *viz.*, to the value *zero* at each point, because of the zero boundary values. This convergence to zero takes place in a regular manner which permits convenient extrapolation.

If the successive difference functions are denoted by $\delta=\psi'-\psi$, δ', δ'', δ''', $\cdots$, the final solution of the problem for this net will be obtained by the addition

$$\Psi=\psi'+\delta'+\delta''+\delta'''+\cdots. \qquad (15)$$

6. EXTRAPOLATION

As discussed in reference 1, it is possible to find certain functions with zero boundary values which, when improved, will be reduced to a constant fraction of themselves without being changed in shape. Such a function is called a characteristic function of the improvement process. The different characteristic functions are denoted by $\phi^{(1)}$, $\phi^{(2)}$, $\phi^{(3)}$, $\cdots$. On improvement these go over to

$$\phi^{(1)\prime}=\lambda_1\phi^{(1)}, \quad \phi^{(2)\prime}=\lambda_2\phi^{(2)}, \quad \text{etc.}$$

The λ values (characteristic values) will all be less than unity, and one of them, say λ_1, will be the largest. It is now possible to make a sort of Fourier expansion of the actual difference function δ in terms of the characteristic functions,[2] i.e.,

$$\delta=a_1\phi^{(1)}+a_2\phi^{(2)}+a_3\phi^{(3)}+\cdots.$$

On improvement this becomes successively

$$\delta'=a_1\lambda_1\phi^{(1)}+a_2\lambda_2\phi^{(2)}+a_3\lambda_3\phi^{(3)}+\cdots,$$
$$\delta''=a_1\lambda_1^2\phi^{(1)}+a_2\lambda_2^2\phi^{(2)}+\cdots,$$
$$\delta'''=a_1\lambda_1^3\phi^{(1)}+a_2\lambda_2^3\phi^{(2)}+\cdots, \quad \text{etc.}$$

Thus after k improvements, the coefficients of the various characteristic functions in δ are decreased by factors λ^k. After a time only the function with the highest characteristic value will remain, and $\delta^{[k]}$ will become a fairly smooth, pillow-shaped, one-signed function decreasing uniformly by a factor λ_1 at each iteration. At this stage one has

$$\delta^{[k]}=c\phi^{(1)},$$
$$\delta^{[k+1]}=\lambda_1 c\phi^{(1)},$$
$$\delta^{[k+2]}=\lambda_1^2 c\phi^{(1)}, \quad \text{etc.}$$

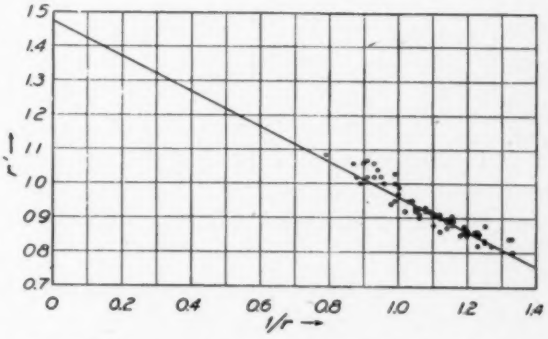


FIG. 3. Extrapolation. Plot of ratios in Eq. (18) for first extrapolation on net shown in Fig. 7. The straight line has the equation $r'=1.475-0.513(1/r)$ so $\lambda_1\approx0.914$, $\lambda_2\approx0.561$, and the right member of Eq. (19) becomes $-13\delta^{[k-1]}+26\delta^{[k]}$.

[2] This statement is not mathematically rigorous, but is adequate for the practical application of the extrapolation scheme. For discussion of this point, see the first paper of reference 1.

so that

$$\delta^{[k]}+\delta^{[k+1]}+\delta^{[k+2]}+\cdots$$
$$=c\phi^{(1)}(1+\lambda_1+\lambda_1^2+\cdots)=\delta^{[k]}/(1-\lambda_1),$$

which furnishes the evaluation of the infinite sum occuring in (15).

This was the method of extrapolation contemplated in reference 1. But is it not necessary to wait for all characteristic functions except $\phi^{(1)}$ to become small before extrapolating. For a fine net (λ values close to unity), it will cut the labor by a large factor if we can extrapolate when *two* of the functions, $\phi^{(1)}$ and $\phi^{(2)}$, remain of importance. Let us suppose that we have improved the difference function k times and that we may write

$$\begin{aligned}\delta^{[k-2]}&=c_1\phi^{(1)}+c_2\phi^{(2)},\\ \delta^{[k-1]}&=\lambda_1c_1\phi^{(1)}+\lambda_2c_2\phi^{(2)},\\ \delta^{[k]}&=\lambda_1^2c_1\phi^{(1)}+\lambda_2^2c_2\phi^{(2)},\quad\text{etc.}\end{aligned}\tag{16}$$

Let us define two successive ratios by

$$r=\delta^{[k-1]}/\delta^{[k-2]};\quad r'=\delta^{[k]}/\delta^{[k-1]}.\tag{17}$$

If $c_1\phi^{(1)}$ and $c_2\phi^{(2)}$ are eliminated from these two ratios, we find that if (16) are satisfied, then

$$r'=(\lambda_1+\lambda_2)-\lambda_1\lambda_2(1/r).\tag{18}$$

Hence we can test whether a net has reached the stage where the differences obey fairly well the relations (16) by computing r and r' at various points and plotting r' against $(1/r)$. If (16) are exactly obeyed, these points will fall on a straight line of slope $(-\lambda_1\lambda_2)$ and intercept $(\lambda_1+\lambda_2)$. This plot will not only tell whether (16) are approximately valid, but will determine the values of λ_1 and λ_2 to use in the extrapolation. Such a plot is given in Fig. 3 by way of illustration.

With characteristic values determined as above for the net as a whole, the values of $c_1\phi^{(1)}$ and $c_2\phi^{(2)}$ at any point are fixed by any two successive δ's of (16). Algebraic manipulation shows that if (16) are obeyed, then

$$\delta^{[k-1]}+\delta^{[k]}+\delta^{[k+1]}+\delta^{[k+2]}+\cdots$$
$$=\delta^{[k-1]}\frac{1-\lambda_1-\lambda_2}{(1-\lambda_1)(1-\lambda_2)}+\frac{\delta^{[k]}}{(1-\lambda_1)(1-\lambda_2)}.\tag{19}$$

Hence the extrapolation proceeds as follows: Suppose that we have computed

$$\begin{aligned}&\psi,\ \psi',\\ &\delta,\ \delta',\ \delta'',\ \cdots,\ \delta^{[k-2]},\ \delta^{[k-1]},\ \delta^{[k]},\end{aligned}\tag{20}$$

and believe that the differences are proceeding regularly enough to try an extrapolation under the assumption of (16). Compute r and r' for all points of the net which give enough significant figures to be useful in λ-value determinations. Make a plot like Fig. 3; from it determine λ_1 and λ_2. Compute the coefficients $(1-\lambda_1-\lambda_2)/(1-\lambda_1)(1-\lambda_2)$ and $1/(1-\lambda_1)(1-\lambda_2)$ in (19). Then compute the extrapolated function value Ψ at each point by adding (see (15))

$$\Psi=\psi'+\delta'+\delta''+\cdots+\delta^{[k-2]}$$
$$+\frac{1-\lambda_1-\lambda_2}{(1-\lambda_1)(1-\lambda_2)}\delta^{[k-1]}+\frac{1}{(1-\lambda_1)(1-\lambda_2)}\delta^{[k]}.\tag{21}$$

Note that while the three last differences of (20) are assumed to obey (16) for the purpose of computing λ_1 and λ_2, only the last two differences $\delta^{[k-1]}$ and $\delta^{[k]}$ need obey (16) for the extrapolation (21) to be valid.

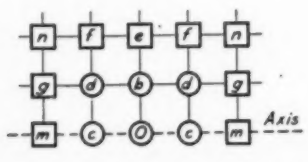


FIG. 4. A nine-block split by the axis.

7. $K=1$, LAPLACE'S EQUATION

In this case we run into no difficulty with negative coefficients under any circumstance. Table I gives the coefficients by which the values at the outside points of Fig. 2 are to be multiplied to obtain the improved value ψ_0 at the center point of the block. If these coefficients are used as multipliers on the computing machine, a check of the computation is furnished by the fact that their sum, after all 12 values have been used, is unity. In addition to these values, it is convenient to have those for a block split by the axis as in Fig. 4. For this block the center point

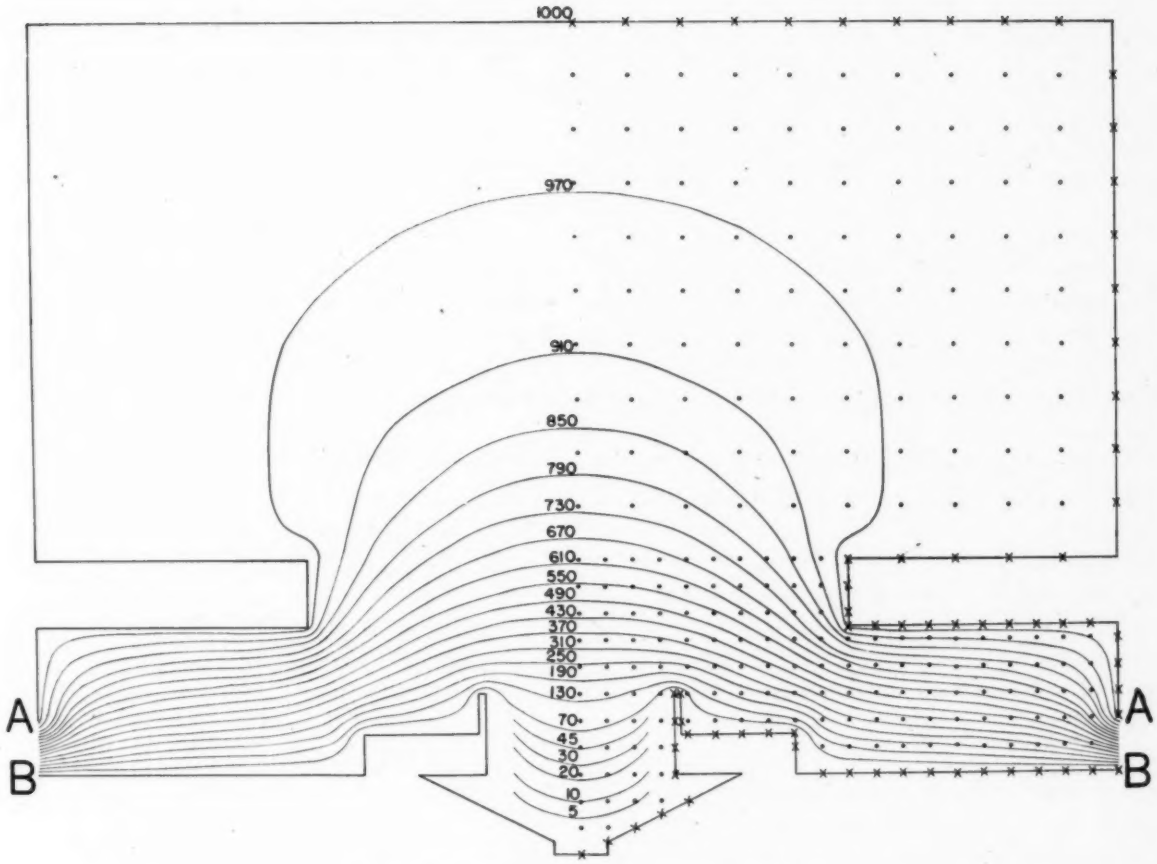


FIG. 5. Equipotential surfaces for an electron lens.

is given by

$$376\psi_0 = 16\psi_m + 40\psi_g + 60\psi_f + 144\psi_e + 60\psi_f + 40\psi_g + 16\psi_m, \quad (22)$$

where the boundary points are taken in order around the block. These coefficients, divided by 376, are listed for $I=0$ in Table I.

As an example of the application of this numerical method we print Fig. 5, which shows the equipotential surfaces for an electron lens in whose properties we were interested. The axis in this figure is vertical. The potential of the upper electrode AA is taken as 1000 volts, that of the lower electrode BB as zero. The net points and boundary points used in the computation are indicated by circles and crosses, respectively. The equipotential surfaces are accurate for the difference equation on these net points; the departure from the solution of the differential equation is probably about one percent over most of the net. We were particularly interested in the degree of penetration of the field into the almost enclosed spaces in the two electrodes.

8. THE TORSION OF AN AXIALLY SYMMETRIC SHAFT

The stresses in a circular shaft of variable diameter subjected to torsion may be expressed[3] in terms of a function $\phi(\rho, z)$ which satisfies the differential equation

$$\frac{\partial^2\phi}{\partial\rho^2} + \frac{\partial^2\phi}{\partial z^2} - \frac{3}{\rho}\frac{\partial\phi}{\partial\rho} = 0. \quad (23)$$

In terms of this function, the stresses have the values:

Normal stresses: $\sigma_\rho = \sigma_z = \sigma_\varphi = 0.$

[3] S. Timoshenko, *Theory of Elasticity* (McGraw-Hill Book Company, Inc., New York, 1934), p. 276 *et seq.*; our definition of ϕ differs by a factor 2π from that of Timoshenko. Physically, $\phi(\rho, z)$ is the torque transmitted across a surface bounded by the circle $\rho=\rho$, $z=z$.

Shears:

$$\tau_{\rho z}=0,$$

$$\tau_{\varphi\rho}=-\frac{1}{2\pi\rho^2}\frac{\partial\phi}{\partial z}, \qquad (24)$$

$$\tau_{\varphi z}=\frac{1}{2\pi\rho^2}\frac{\partial\phi}{\partial\rho}.$$

Boundary conditions require that $\phi=0$ along the axis or on the inner boundary if the shaft is partially or entirely hollow; $\phi=M$, the torque, on the outer boundary. In a diametral (ρ, z) plane, lines of constant ϕ are in the direction of the maximum shear.

For a straight circular shaft of radius a, the function ϕ is $M\rho^4/a^4$. The only component of stress is $\tau_{\varphi z}=2M\rho/\pi a^4$.

Since $\tau_{\varphi z}$ must vanish at $\rho=0$, we see from (24) that for any shaft ϕ must vary like ρ^4 near the axis. A function varying this rapidly is inconvenient to handle by numerical methods; it is simpler to work with the function

$$\Psi=\phi/\rho^4 \qquad (25)$$

which will start like $A+B\rho^2+\cdots$ (where A and B are functions of z) near the axis.

The function Ψ will satisfy the differential equation

$$\frac{\partial^2\Psi}{\partial\rho^2}+\frac{\partial^2\Psi}{\partial z^2}+\frac{5}{\rho}\frac{\partial\Psi}{\partial\rho}=0. \qquad (26)$$

Boundary values for Ψ are determined from those of ϕ according to (25); along the axis Ψ does not vanish but $\partial\Psi/\partial\rho=0$, and in fact, if Ψ be expanded in a power series, only even powers of ρ may occur.

For a straight circular shaft,

$$\Psi=\text{const.}=M/a^4.$$

For a grooved or filleted shaft with long straight cylindrical end sections, Ψ is constant except in the vicinity of the groove or fillet where the stresses depart from the circular-shaft values. For this reason, we call Ψ the *stress-concentration function.*

To complete this discussion of elastic theory, we note that if $\Theta(\rho, z)$ represents the angle of twist in radians, at any point, this function is related to the stress function $\phi(\rho, z)$ as follows:[3]

$$\frac{\partial\Theta}{\partial\rho}=-\frac{1}{2\pi G\rho^3}\frac{\partial\phi}{\partial z},$$

$$\frac{\partial\Theta}{\partial z}=\frac{1}{2\pi G\rho^3}\frac{\partial\phi}{\partial\rho},$$

where G is the modulus of elasticity in shear. Hence, for a straight circular shaft of radius a, Θ is $2Mz/\pi a^4G$. From the above relations, it follows that in the ρ, z plane, lines of constant Θ are orthogonal to lines of constant ϕ.

9. $K=5$, THE STRESS-CONCENTRATION FUNCTION

We consider next the application of the numerical methods of Sections 2–6 to the solution of Eq. (26).

Coefficients in the simple improvement formulas (4) and (7) become negative at $I=1$ and $I=2$; special formulas for these points are therefore needed.

For $I=2$, formula (4) gives (cf. Fig. 1)

$$16\psi_0=4\psi_1+4\psi_2+9\psi_3-\psi_4. \qquad (27)$$

This formula is not adapted to the iteration process because of the negative coefficient. It was derived by passing a quadric function through points (0), (1), (2), (3), and (4) which satisfied the differential equation at (0). An alternative formula (see Appendix 1) is obtained by passing a fourth-order polynomial *solution* of the differential equation through the five points. For such a solution,

$$4\psi_0=10\psi_1+10\psi_2+9\psi_3-25\psi_4. \qquad (28)$$

Use of such fourth-order functions is not advisable in numerical work except on extremely fine nets. But the following linear combination of (27) and (28) eliminates the negative coefficient of ψ_4, and involves much more of (27) than (28):

$$22\psi_0=5\psi_1+5\psi_2+12\psi_3 \quad (K=5,\ I=2). \qquad (29)$$

It is proposed to use this as the ordinary improvement formula at $I=2$. Similar considerations will not give a diagonal formula for $I=2$ which is free from negative coefficients. Hence,

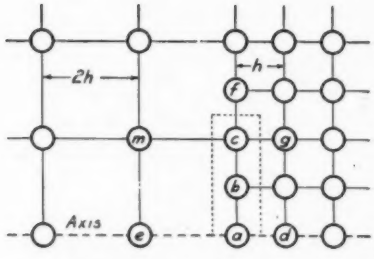


FIG. 6. Change of net spacing near axis.

no diagonal improvement formula is available for $K=5$, $I=2$.

For $I=1$, $K=5$, the simplest and most satisfactory improvement formula is

$$4\psi_0=\psi_3+3\psi_4 \quad (K=5,\ I=1) \tag{30}$$

(cf. Fig. 1, where point (4) is now on the axis). The formula can be obtained by the method used for (29), but more simply, is an approximation to any function that is even in ρ. Again, no satisfactory diagonal improvement formula is available for $I=1$.

Change of net spacing near axis: For points b and c of Fig. 6, the previous formulas (9) and (5) will involve negative coefficients. For b, formula (30)

$$4\psi_b=\psi_c+3\psi_a \tag{31}$$

is available. For c, if a cubic *solution* is passed through b, c, f, g, m, it is found that

$$8\psi_c=5\psi_b+3\psi_f, \tag{32}$$

which is simple and satisfactory. For a, (13) gives

$$39\psi_a=36\psi_b+2\psi_d+\psi_e. \tag{33}$$

It is seen that points a, b, c may be improved simultaneously as a block, thus speeding up the convergence. New values for a, b, c, are obtained from those at d, e, f by using successively the following three formulas, which are derived from (31)–(33):

$$\begin{aligned} 7\psi_a&=2\psi_d+\psi_e+4\psi_f,\\ 9\psi_b&=8\psi_a+\psi_f,\\ 8\psi_c&=5\psi_b+3\psi_f. \end{aligned} \tag{34}$$

Nine-blocks: The regular nine-block procedure sketched in Section 4 can be used only where the center point has $I\geqslant 4$. Table II gives the coefficients used for improvement of the center point. The formulas for the axis nine-block (Fig. 4) may be obtained by use of (30) for points b and d. The center point is given by

$$23\psi_0=15\psi_e+3\psi_f+3\psi_f+\psi_m+\psi_m.$$

The other points are then obtained successively from

$$\begin{aligned} 5\psi_c&=3\psi_f+\psi_0+\psi_m,\\ 4\psi_b&=\psi_e+3\psi_0, \quad 4\psi_d=\psi_f+3\psi_c. \end{aligned}$$

Blocks centered at $I=2$ or 3 are not available because diagonal formulas are needed for the corner points. Hence it is desirable to arrange blocks along the axis and around $I=4$, 7, etc. (cf. Fig. 7). This leaves the row at $I=2$ to be computed from (29). Since the points at $I=1$ do not enter (29), the values at $I=0$ and 1 are frequently absent from the iteration formulas and may be left for computation after all other values have converged.

10. STRESS CONCENTRATION IN A GROOVED SHAFT

In order to illustrate these numerical methods and to show how the various quantities of physical interest may be obtained from the stress-concentration function Ψ, a complete solution of the torsion problem for the particular grooved shaft shown in Fig. 7 was carried through.

While we are interested in a shaft of this geometrical shape with arbitrary radius R, torque M, and shear modulus G, it is simplest to carry through the numerical computation for the case of unit radius, unit torque, and unit shear modulus, and then to multiply the results by factors which are determined by the physical dimensions of the quantities concerned, i.e., shear values by M/R^3 and angles of twist by M/R^3G.

Stress-Concentration Function. For the shaft of unit radius and unit torque, $\phi=1$ on the boundary and hence the stress-concentration function Ψ equals $1/\rho^4$ on the boundary and approaches 1 on the straight part of the shaft. It is simpler to compute the function $(\Psi-1)$, which satisfies the same differential equation as Ψ and approaches 0 on the straight parts of the shaft. Figure 8 shows contours of this function. The points used in the finest net and the division into blocks are shown in Fig. 7. The values of Fig. 8 are believed

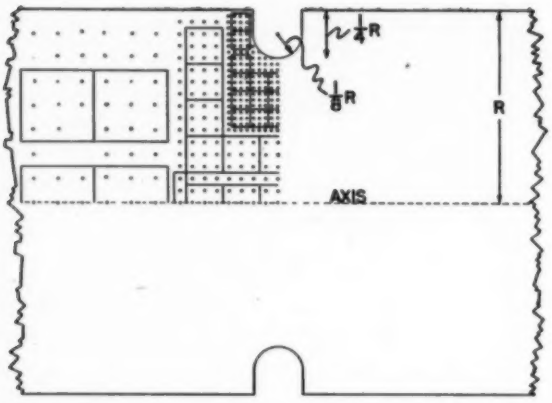


FIG. 7. Grooved shaft showing arrangement of net points and nine-blocks used in numerical stress computation.

to be sufficiently accurate to permit computation of stresses to better than 1 percent accuracy.

The stress function ϕ for the shaft of unit radius and unit torque is plotted in Fig. 9. Contours are given for $\phi^{\frac{1}{2}}$, which is $\rho\Psi^{\frac{1}{2}}$.

The twist angle Θ is shown in Fig. 10. The values plotted are those of $2\pi\Theta$ for a shaft of unit radius, unit torque, and unit shear modulus. These values were computed from the formula

$$2\pi\Theta(\rho,z)=2\pi\Theta(\rho,z_0)+\int_{z_0}^{z}\left(4\Psi+\rho\frac{\partial\Psi}{\partial\rho}\right)dz. \quad (35a)$$

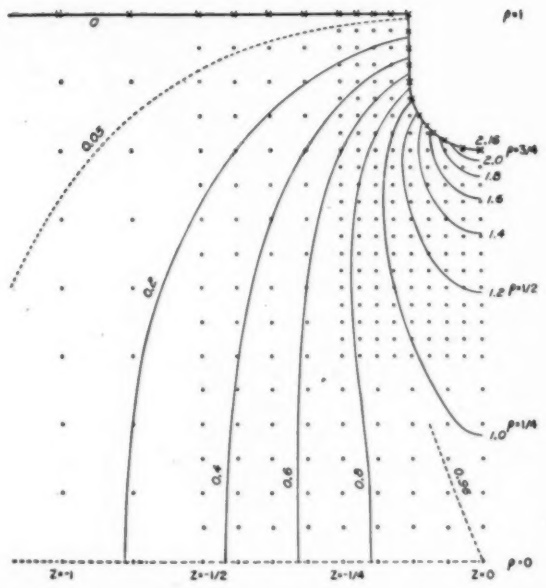


FIG. 8. The stress-concentration function $(\Psi-1)$ for the shaft of unit radius and unit torque.

The value of Θ was set equal to zero at $z_0=-1\frac{1}{4}$ (measuring from the plane of symmetry at the groove) where the function $\Psi-1$ essentially vanishes and the twist is like that of a straight circular shaft. $\partial\Psi/\partial\rho$ was evaluated by the simple parabolic formula

$$\partial\Psi(\rho,z)/\partial\rho=[\Psi(\rho+h,z)-\Psi(\rho-h,z)]/2h$$

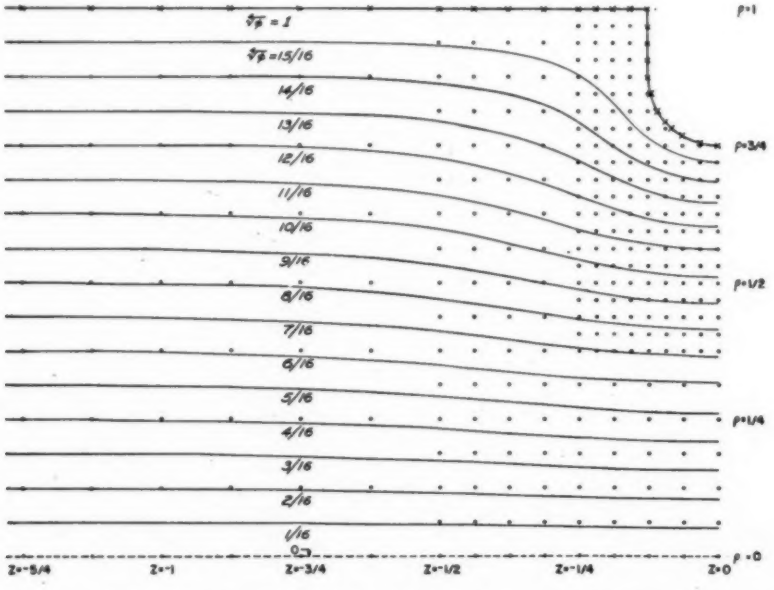


FIG. 9. The stress function ϕ for the shaft of unit radius and unit torque. Lines of constant ϕ show the directions of the shear vector in the diametral plane.

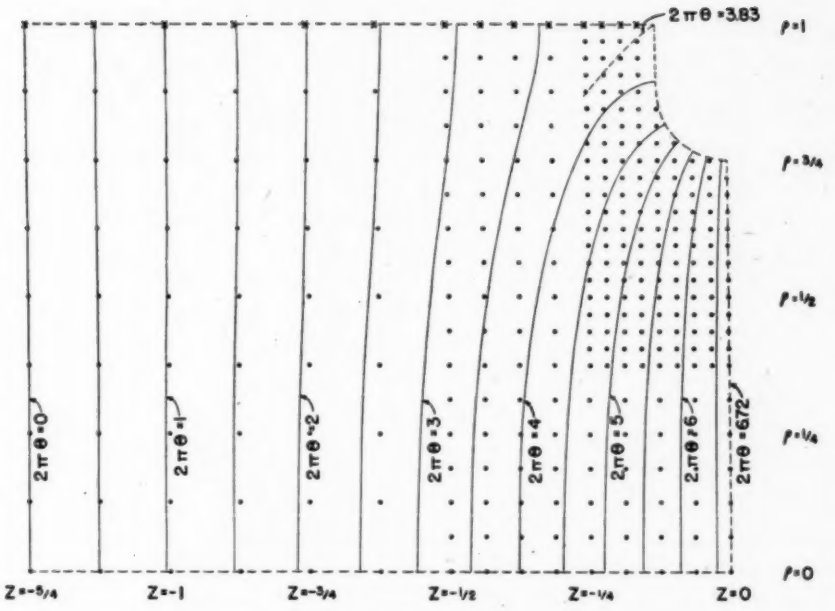


FIG. 10. The twist angle Θ in radians for the shaft of unit radius, unit torque, and unit shear modulus. For the shaft shown in Fig. 7, these values of Θ are to be multiplied by M/R^3G. Lines of constant Θ are orthogonal to those of constant ϕ shown in Fig. 9, but there is no simple relation between the spacing of the lines on the two plots.

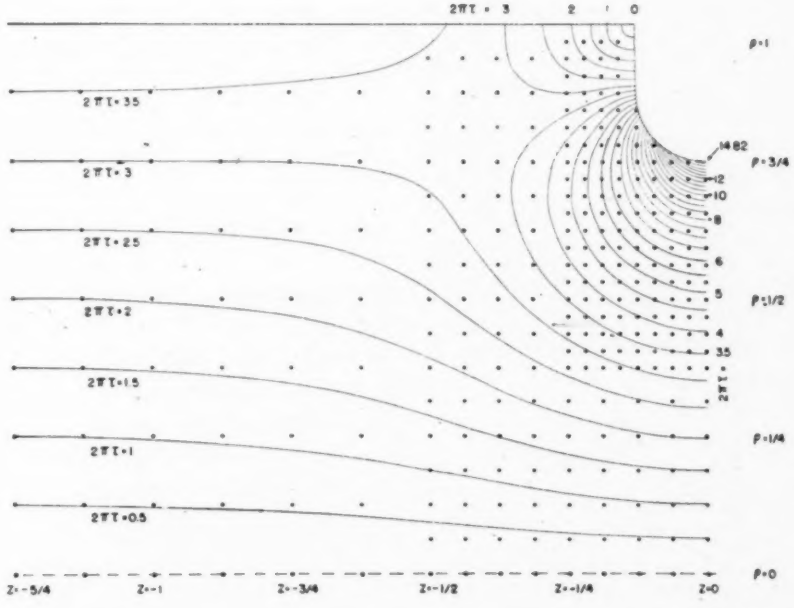


FIG. 11. Magnitude, $\tau=(\tau_{\varphi\rho}^2+\tau_{\varphi z}^2)^{\frac{1}{2}}$, of the shear vector in the axial plane for the shaft of unit radius with unit torque. The curves of Fig. 9 show the direction of this vector. For the shaft shown in Fig. 7, these values of τ are to be multiplied by M/R^3.

and then the integration was carried out by simple parabolic rules. A check was furnished by the constancy of angle of twist at the plane of symmetry at the groove. Intermediate points were filled in and further checks obtained from the formula

$$2\pi\Theta(\rho, z) = 2\pi\Theta(0, z) - \int_0^\rho \rho \frac{\partial \Psi}{\partial z}\, d\rho. \quad (35b)$$

It is pointed out that much greater numerical accuracy is obtained in both the angle of twist and in the stresses discussed below by computing them from Ψ rather than from ϕ.

Figure 10 gives contours of constant twist angle. For a straight shaft without a groove, the contours shown would be uniformly spaced at intervals of $\frac{1}{8}$ in z, and the twist at $z=0$ would be $5/2\pi$ radians instead of $6.72/2\pi$ radians for the grooved shaft. The extra twist occasioned by the presence of the groove is thus $1.72/2\pi$ in half the shaft length, or $3.44/2\pi = 0.547$ radian in the whole shaft. For a shaft of radius R, torque M, and shear modulus G, the extra twist introduced by this shape of groove is thus

$$\text{Extra twist} = 0.547 M/R^3 G \text{ radians},$$

which is equivalent to the twist in a length $0.86R$ of a straight circular shaft of radius R. Note that the rate of twist in the section of full diameter decreases as the notch is approached and that this deficiency in twist is made up in the notch in addition to the extra twist given above. One cannot, therefore, determine the extra twist by attaching a torsiometer to the edges of the notch; the value measured will be too great.[4] In our shaft this deficiency amounts to $0.67/2\pi$ radian on the full-diameter part of the shaft near each side of the notch.

Finally, *values of the maximum shear*,

$$\tau = (\tau_{\varphi\rho}^2 + \tau_{\varphi z}^2)^{\frac{1}{2}}$$

are given in Fig. 11. Contours are given for constant values of $2\pi\tau$ for the shaft of unit radius and unit torque. Such values of τ are to be multiplied by M/R^3 in the general case. This plot gives the magnitude of the shear vector in the axial plane, while the curves of Fig. 9 show the direction of this vector.

Values of $2\pi\tau$ were obtained by first computing the components from the formulas

$$\begin{aligned} 2\pi\tau_{\varphi\rho} &= -\rho^2 \partial\Psi/\partial z, \\ 2\pi\tau_{\varphi z} &= \rho(4\Psi + \rho\partial\Psi/\partial\rho). \end{aligned} \quad (36)$$

It will be noticed that these formulas involve the same functions of Ψ as in the formulas (35) for twist angle.

The high stress-concentration at the root of the groove is at once apparent from Fig. 11. The maximum stress is

$$2\pi\tau_{\max} = 14.82 M/R^3.$$

This is 3.70 times the maximum stress

$$(2\pi\tau_{\max} = 4M/R^3)$$

for a straight circular shaft of radius R, and is 1.56 times the maximum stress in a straight circular shaft of radius $\frac{3}{4}R$. This last factor, which compares the maximum stress in the grooved shaft to that in a straight shaft of diameter equal to the root diameter is usually called the *stress-concentration factor* k, so for this shaft $k=1.56$. This value is believed to be accurate to within 1 percent.

Our value, 1.56, for the stress-concentration factor may be compared with the experimental value 1.45 read from McGivern's curves of Fig. 4, Bulletin No. 62.[4] The agreement is within the experimental accuracy of McGivern's "Plastic-Flow Method." It may also be compared with Neuber's theory[5] for a "deep external notch." Neuber gives the formula

$$k = \tfrac{3}{4}[1+(s+1)^{\frac{1}{2}}]^2/[1+2(s+1)^{\frac{1}{2}}],$$

where s is the ratio of shaft radius at the root of the groove to the radius of curvature of the

[4] This point seems not to have been taken into consideration by J. G. McGivern, Eng. Expt. Station of the State College of Washington, Eng. Bull. No. 58 and 62 (1939), who has made detailed experimental measurements of the torsion of grooved and filleted shafts. We gather from these papers that he placed his twist measuring gauges directly at the edges of the fillets and grooves and thus obtained values of extra twist which were too high. However, the extra twist determined from McGivern's Fig. 5, Bulletin 62, for our particular groove seems to be about twice our computed value, which is a considerably greater discrepancy than can be accounted for by the above comment, but is not inconsistent with the scatter of McGivern's experimental data.

[5] H. Neuber, *Kerbspannungslehre* (Verlagsbuchhandlung, Julius Springer, Berlin, 1937); translated as *Theory of Notch Stresses*, David Taylor Model Basin Trans. 74, Washington (1945).

groove at the root. In our case $s=\frac{3}{4}R/\frac{1}{8}R=6$. This formula is exact for a hyperbolic notch in an otherwise infinite shaft, and is an approximation to any shape of notch deep enough so that the resulting stress-concentration extends significantly throughout the diameter of the shaft at the notch. The value $s=6$ in Neuber's formula gives $k=1.58$, in excellent agreement with our value of 1.56.

We take pleasure in thanking Mr. B. M. Shepard of the Naval Ordnance Laboratory for assistance in the preparation of the figures for the grooved-shaft problem.

APPENDIX 1. IMPROVEMENT FORMULAS

Consider Fig. 1 and the various cuts in Section 3 showing a point (0) surrounded by four or five neighboring points. The improvement formulas (4)–(11) can be considered as derived in the following way: Let the coordinates of (0) be $z=0$, $\rho=\rho_0$. Write the most general quadric function passing through ψ_0 in the form

$$\psi=\psi_0+\alpha_1 z+\alpha_2(\rho-\rho_0)+\beta_1 z^2+\beta_2 z(\rho-\rho_0)+\beta_3(\rho-\rho_0)^2.$$

Now choose the coefficients so that this quadric passes through the neighboring four or five function values ψ_1 to ψ_4 or ψ_5; e.g., for the regular point of Fig. 1, the quadric function passing through ψ_0, ψ_1, ψ_2, ψ_3, ψ_4 is

$$\begin{aligned}\psi=\psi_0&+(\psi_1-\psi_2)z/2h+(\psi_3-\psi_4)(\rho-\rho_0)/2h\\&+(\psi_1+\psi_2-2\psi_0)z^2/2h^2\\&+(\psi_3+\psi_4-2\psi_0)(\rho-\rho_0)^2/2h^2+\beta_2 z(\rho-\rho_0)\end{aligned}$$

with β_2 arbitrary. Finally, substitute this function in the differential equation (1) and then set $z=0$, $\rho=\rho_0$ to obtain a relation between ψ_0, ψ_1, ψ_2, ψ_3, ψ_4 which is just the improvement formula (4). This procedure for deriving improvement formulas for various arrangements of points can be formally simplified but is stated in this way to show exactly what is done. *A general quadric is passed through the function values and then required to satisfy the differential equation at just the point to be improved.*

An alternative procedure for deriving improvement formulas is to start with a quantic that satisfies the differential equation (1) *everywhere* and then make it pass through the points of interest in the neighborhood of ψ_0. If we are to do this for the points of Fig. 1, a fourth-order polynomial (quartic surface) must be used. The most general quartic *solution* of the differential equation (1) is

$$\begin{aligned}\psi=A+Bz&-(1+K)Cz^2+C\rho^2-\tfrac{1}{3}(1+K)Dz^3+Dz\rho^2\\&+\tfrac{1}{3}(1+K)(3+K)Ez^4-(6+2K)Ez^2\rho^2+E\rho^4,\end{aligned}$$

for arbitrary A, B, C, D, E. Note that this function is automatically even in ρ. If now we attempt to put this surface through the four values ψ_1, ψ_2, ψ_3, ψ_4 of Fig. 1 in order to determine ψ_0, it is found that we must have $E\neq 0$—we cannot put a cubic solution through these values but must use a full quartic. The quartic solution through ψ_1, ψ_2, ψ_3, ψ_4 of Fig. 1 determines a unique value of ψ_0 and gives the improvement formula

$$\begin{aligned}[96I^3-4K^3I-40KI-48I]\psi_0&=[24I^3-6I](\psi_1+\psi_2)\\&+[24I^3+12KI^2-2K^2I-20KI-18I+K^2+7K+6]\psi_3\\&+[24I^3-12KI^2-2K^2I-20KI-18I-K^2-7K-6]\psi_4.\end{aligned}$$

This is a rather complex formula which is probably not so well suited for most numerical work as the simple formula based on a quadric function. Unless the net is fine enough and the function smooth enough so that fourth differences are regular, the use of a quartic function for approximation is apt to introduce rather than to smooth out irregularities. In such numerical work as we are describing it is best to retain simple quadric approximating functions unless one has occasion to reach the point of extreme refinement of net and accuracy.

It is noted that Eq. (28) is an example of the above quartic formula.

APPENDIX 2. COEFFICIENTS FOR IMPROVEMENT OF NINE-BLOCK CENTER POINT

As a matter of record, we give below the complete expressions, for arbitrary I and K, for the nine-block coefficients of which Tables I and II are special cases. If the improvement formula for the center point of Fig. 2 is written in the form

$$\begin{aligned}C_0\psi_0=C_c\psi_c+C_E\psi_E+C_f\Sigma\psi_f+C_F\Sigma\psi_F\\+C_g\Sigma\psi_g+C_G\Sigma\psi_G+C_m\Sigma\psi_m,\end{aligned}$$

where Σ indicates summation of the values at the two points having the same label, the general expressions for the C's are

$$\begin{aligned}C_0&=2I\{(512I^4+64I^2(K^2-2K-16)\\&\qquad+(K^4-4K^3-60K^2+128K+512)\},\\C_c&=128I^5+128I^4K+4I^3(9K^2-18K-64)\\&\quad+4I^2K(K^2-10K-48)+I(K^4-4K^3-32K^2\\&\qquad+72K+128)+K(-K^3+36K+64),\\C_f&=16\{4I^5+4I^4K+I^3(K^2-2K-8)+I^2(-K^2-6K)\\&\qquad+I(-K^3+2K+4)+(K^3+2K)\},\\C_g&=32\{2I^5+I^4K-4I^3-2I^2K+2I+K\},\\C_m&=4I\{32I^4+I^2(-K^2+2K-64)+(K^3-2K+32)\},\end{aligned}$$

with C_E obtained from C_c, C_F from C_f, and C_G from C_g by replacing I by $-I$ and then changing the sign of the whole expression, i.e., $C_E(I,K)=-C_c(-I,K)$, etc.

Note on a Boundary Value Problem of Reissner and Sagoci

IAN N. SNEDDON
The Department of Natural Philosophy, the University of Glasgow, Scotland
(Received June 18, 1946)

1. INTRODUCTION

IN their investigation of the torsional oscillations produced in a semi-infinite, homogeneous, isotropic medium by a periodic shear stress applied in an axially symmetrical manner to a circular area of the surface of the medium, Reissner and Sagoci[1] begin by considering a boundary value problem in the theory of elastic equilibrium. The distribution of stress in the interior of a semi-infinite elastic medium is determined when a load is applied to the surface by means of a rigid disk; the torsional displacement is prescribed immediately under the disk, and it is assumed that that part of the boundary which lies beyond the edge of the disk is free from stress. The solution of this mixed boundary value problem is obtained by the introduction of a certain system of oblate spheroidal coordinates.

In the introduction to their paper Reissner and Sagoci point out that the use of Fourier-Bessel methods only reduces the problem to a problem in integral equations which is then reduced to the solution of an infinite number of linear algebraic equations in an infinite number of unknowns. The object of this note is in part to recall that the problem may be reduced to the solution of a pair of dual integral equations.[2] In the static case these dual integral equations reduce to a type which arises in a number of mixed boundary value problems in the mathematical theory of elasticity,[3] and their solution is known. In the present note the dual integral equations are set up by the use of the Hankel transform theorem. It is then shown that the solution for the static case can be derived in this way; though the dynamic problem is not solved a possible approach to it is sketched in Section 3.

[1] E. Reissner and H. F. Sagoci, J. App. Phys. **15**, 652 (1944); H. F. Sagoci, J. App. Phys. **15**, 655 (1944).
[2] Such a reduction was first given by E. Reissner in a paper which appeared in Ing. Archiv **8**, 229–45 (1937).
[3] See, for example: J. W. Harding and I. N. Sneddon, Proc. Camb. Phil. Soc. **41**, 12 (1945); I. N. Sneddon, Proc. Roy. Soc. A, to be published; I. N. Sneddon and H. A. Elliott, Quart. App. Math., to be published.

2. FORMULATION OF THE PROBLEM

We wish to determine the components of stress and of the displacement vector in the interior of the semi-infinite, homogeneous, isotropic, elastic solid $z \geqslant 0$ when a circular area ($r \leqslant r_0$) of the surface is forced to rotate through an angle Φ about an axis which is normal to the undeformed surface of the medium. It is assumed that the region of the surface lying out with the circle $r \leqslant r_0$ is free from stress. It has been shown by Reissner[2] that, in this case, only the circumferential component v of the displacement vector is different from zero, and that all the components of stress vanish identically except $\tau_{z\theta}$ and $\tau_{r\theta}$ which are given by the relations

$$\tau_{z\theta}=\mu\frac{\partial v}{\partial z}, \quad \tau_{r\theta}=\mu\left(\frac{\partial v}{\partial r}-\frac{v}{r}\right), \tag{1}$$

where, in the absence of internal damping, v satisfies the partial differential equation

$$\frac{\partial^2 v}{\partial r^2}+\frac{1}{r}\frac{\partial v}{\partial r}-\frac{v}{r^2}+\frac{\partial^2 v}{\partial z^2}=\frac{1}{c^2}\frac{\partial^2 v}{\partial t^2}, \tag{2}$$

where $c=(\mu/\rho)^{\frac{1}{2}}$. The boundary conditions of the problem are

$$z=0, \quad v=f(r,t), \quad r\leqslant r_0, \tag{3}$$

$$z=0, \quad \tau_{z\theta}=0, \quad r\geqslant r_0. \tag{4}$$

In the case considered by Reissner and Sagoci

$$f(r,t)=\Phi(t)\cdot r. \tag{5}$$

3. REDUCTION OF THE PARTIAL DIFFERENTIAL EQUATION BY MEANS OF THE HANKEL TRANSFORM

To solve the partial differential Eq. (2) we introduce the Hankel transform

$$\bar{v}=\int_0^\infty rvJ_1(\xi r)dr,$$

of the circumferential component v of the displacement vector. Hence if we multiply both

sides of Eq. (2) by $rJ_1(\xi r)$ and integrate with respect to r from 0 to ∞ we have the differential equation

$$\left(\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial z^2}+\xi^2\right)\bar{v}=0 \qquad (6)$$

for the determination of the function $\bar{v}(z, t, \xi)$. Once this solution is known we obtain the expression for the displacement v by means of the Hankel inversion theorem[4]

$$v=\int_0^\infty \xi\bar{v}(z, t, \xi)J_1(\xi r)d\xi \qquad (7)$$

from which we have for the stress component $\tau_{z\theta}$

$$\tau_{z\theta}=\mu\int_0^\infty \xi\frac{\partial\bar{v}(z, t, \xi)}{\partial z}J_1(\xi r)d\xi. \qquad (8)$$

The arbitrary functions introduced in the solution of the partial differential Eq. (7) are determined by the boundary conditions. The solution must be such that the displacement and both components of stress tend to zero as $z\to\infty$; when $z=0$ the conditions (3) and (4) must be satisfied. Substituting these relations into the expressions (7) and (8) for v and $\bar{v}$ we obtain the dual integral equations

$$\int_0^\infty \xi\bar{v}(0, t, \xi)J_1(\xi r)d\xi=f(r, t), \quad r\leqq r_0, \qquad (9)$$

$$\int_0^\infty \xi\frac{\partial\bar{v}(0, t, \xi)}{\partial z}J_1(\xi r)d\xi=0, \quad r\geq r_0 \qquad (10)$$

for the determination of the remaining arbitrary part of the function $\bar{v}(z, t, \xi)$. These equations are equivalent to Eqs. (129) and (130) of Reissner's paper.[2]

4. SOLUTION FOR THE STATIC CASE

In the static case the function $\Phi(t)$ of Eq. (5) is a constant, or, in the general case, the function $f(r, t)$ of Eq. (3) is a function of r alone—say $f(r)$. We may therefore take $\partial/\partial t$ to be identically zero in Eq. (7) so that it reduces to an ordinary differential equation with solution

$$\bar{v}=A(\xi)e^{-\xi z}+B(\xi)e^{\xi z}. \qquad (11)$$

Since v, and hence $\bar{v}$, tends to zero as $z\to\infty$ we must take $B(\xi)$ to be identically zero in Eq. (11). Substituting this value for v into the Eqs. (10) and (9) we obtain the dual integral equations

$$\int_0^\infty \xi A(\xi)J_1(\xi r)d\xi=f(r), \quad r\leqq r_0 \qquad (12)$$

$$\int_0^\infty \xi^2 A(\xi)J_1(\xi r)d\xi=0 \qquad r\geq r_0 \qquad (13)$$

which are Reissner's Eqs. (129) and (130) with $k=0$. Writing $r=\rho r_0$, $\xi=\eta/r_0$, $\eta^2A(\eta)=F(\eta)a^3$, $f(r)=ag(\rho)$, we see that these equations reduce to

$$\int_0^\infty \eta^{-1}F(\eta)J_1(\eta\rho)d\eta=g(\rho), \quad \rho<1,$$

$$\int_0^\infty F(\eta)J_1(\eta\rho)d\eta=0, \qquad \rho>1.$$

Dual integral equations of this type have been considered by Titchmarsh[4] and Busbridge.[5] By making use of the theory of Mellin transforms and known theorems in the calculus of residues Titchmarsh obtained the solution of a generalization of this pair of dual integral equations in the form of a contour integral. The reduction to real form was effected by Busbridge and covers the case arising here; the solution turns out to be

$$F(\eta)=\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\eta\left\{\eta^{\frac{1}{2}}J_{\frac{1}{2}}(\eta)\int_0^1 y^2(1-y^2)^{-\frac{1}{2}}g(y)dy\right.$$

$$\left.+\int_0^1 u^2(1-u^2)^{-\frac{1}{2}}du\int_0^1 g(yu)(\eta y)^{\frac{1}{2}}J_{3/2}(\eta y)dy\right\}.$$

In the case considered by Reissner and Sagoci

$$g(\rho)=\Phi\rho,$$

so that

$$F(\eta)=\frac{4\Phi}{\pi}\left(\frac{\sin\eta}{\eta}-\cos\eta\right)$$

giving finally

$$v=\frac{4\Phi a}{\pi}\int_0^\infty\left(\frac{\sin\eta}{\eta^2}-\frac{\cos\eta}{\eta}\right)e^{-\xi\eta}J_1(\rho\eta)d\eta,$$

[4] E. C. Titchmarsh, *An Introduction to the Theory of Fourier Integral* (Oxford, 1937), p. 240.

[5] I. W. Busbridge, Proc. London Math. Soc. [2], **44**, 115 (1938).

where $\zeta = z/r_0$. Now it is easily shown that[6]

$$I_1 = \int_0^\infty \frac{\sin\eta}{\eta^2} e^{-\zeta\eta} J_1(\rho\eta)\, d\eta = \frac{\lambda R \sin(\theta+\phi) - \lambda^2 \sin 2\theta}{2\rho} + \tfrac{1}{2}\rho \tan^{-1} \frac{R\sin\phi + \lambda\sin\theta}{R\cos\phi + \lambda\cos\theta},$$

$$I_2 = \int_0^\infty \frac{\cos\eta}{\eta} e^{-\zeta\eta} J_1(\rho\eta)\, d\eta = \frac{1}{\rho}(R\cos\phi - \zeta),$$

$$I_3 = \int_0^\infty \frac{\sin\eta}{\eta} e^{-\zeta\eta} J_1(\rho\eta)\, d\eta = \frac{1}{\rho}(1 - R\sin\phi),$$

$$I_4 = \int_0^\infty \cos\eta\, e^{-\zeta\eta} J_1(\rho\eta)\, d\eta = \frac{1}{\partial}\left\{1 - \frac{\lambda}{R}\cos(\theta-\phi)\right\},$$

where the quantities λ, R, θ, and ϕ are defined by the relations

$$\lambda^2 = 1+\zeta^2, \qquad \zeta\tan\theta = 1,$$

$$R^4 = (\rho^2+\zeta^2-1)^2 + 4\zeta^2, \quad 2\zeta\cot 2\phi = \rho^2+\zeta^2-1.$$

Hence we have the formulae

$$v = \frac{2\Phi a}{\pi\rho}\left[\lambda R\sin(\theta+\phi) - \lambda^2\sin 2\theta + \rho^2 \tan^{-1}\frac{R\sin\phi+\lambda\sin\theta}{R\cos\phi+\lambda\cos\theta} - 2R\cos\phi + 2\zeta\right],$$

$$\tau_{z\theta} = \frac{4\Phi\mu R}{\pi\rho}\left[\sin\phi - \frac{\lambda}{R}\cos(\theta-\phi)\right],$$

for the determination of v to $\tau_{z\theta}$ at any point in the interior of the elastic medium; a similar expression may be derived for the other shear stress $\tau_{r\theta}$. When $z=0$ and $\rho>1$,

$$I_1 = \frac{1}{2\rho}\left[(\rho^2-1)^{\frac{1}{2}} + \rho^2\left\{\frac{\pi}{2} - \tan^{-1}(\rho^2-1)^{\frac{1}{2}}\right\}\right],$$

$$I_2 = (\rho^2-1)^{\frac{1}{2}}/\rho,$$

and when $z=0$, $\rho<1$,

$$I_3 = \frac{1}{\rho}[1-(1-\rho^2)^{\frac{1}{2}}], \quad I_4 = \frac{1}{\rho}[1-(1-\rho^2)^{-\frac{1}{2}}],$$

giving for v and $\tau_{z\theta}$ when $z=0$,

$$v = \Phi a\rho\left[\left\{1 - \frac{2}{\pi}\tan^{-1}(\rho^2-1)^{\frac{1}{2}}\right\} - \frac{2}{\pi}(1-1/\rho^2)^{\frac{1}{2}}\right], \qquad \rho>1,$$

$$\tau_{z\theta} = -\frac{4\Phi\mu}{\pi}(1/\rho^2-1)^{-\frac{1}{2}}, \quad \rho<1$$

in agreement with the expressions found by Reissner and Sagoci.[7]

[6] See, for instance, I. N. Sneddon, Proc. Camb. Phil. Soc. **42**, 27 (1946).

[7] Reference 6, Eqs. (23) and (24).

Erratum: Theory of Filler Reinforcement

[J. App. Phys. **16**, 55 (1945)]

JANE M. DEWEY
United States Rubber Company, Passaic, New Jersey

THE last line of Eq. (3) should read

$$-\frac{1}{\gamma-1}\left[\frac{\partial P_{\gamma-2}^{|m|}}{\partial\theta} i_\theta + \frac{im}{\sin\theta} P_{\gamma-2}^{|m|} i_\phi\right].$$

In Eq. (5)

$$M = -\tfrac{1}{3}(A+B-2C)P_2(\cos\theta) + \tfrac{1}{12}(A-B)P_2^2(\cos\theta)(e^{2i\phi}+e^{-2i\phi}).$$

In Eq. (6)

$$b_{-2} = (3\lambda_1+2\lambda_2-3\lambda_1'-2\lambda_2')(4\lambda_2+3\lambda_1'+2\lambda_2')^{-1}R^3,$$
$$c_{-2} = -5(3\lambda_1+5\lambda_2)DR^3,$$
$$c_{-4} = 9(\lambda_1+\lambda_2)DR^5,$$
$$a_1 = (A+B+C)(\lambda_1+2\lambda_2)(4\lambda_2+3\lambda_1'+2\lambda_2')^{-1}.$$